DOCUMENT RESUME

ED 273 297 IR 051 622

AUTHOR Lancaster, F. W.
TITLE Thesaurus Construction and Use. A Condensed Course.
INSTITUTION United Nations Educational, Scientific and Cultural Organization, Paris (France). General Information Programme.
REPORT NO PGI-85/WS/11
PUB DATE Mar 85
NOTE 96p.
PUB TYPE Guides - Classroom Use - Guides (For Teachers) (052) -- Audiovisual Materials (100)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Autoinstructional Aids; Higher Education; *Instructional Materials; *Lexicography; *Subject Index Terms; Teaching Guides; *Thesauri; Visual Aids; *Vocabulary
IDENTIFIERS *Natural Language

ABSTRACT

Prepared as a guide for the use of an instructor, this document contains two principal components: 84 "exhibits," which could be converted into slides or transparencies, and text, which explains and amplifies the exhibits. The text discusses 14 subject areas: (1) Purpose of Vocabulary Control; (2) Major Components of a Controlled Vocabulary; (3) Gathering Terms; (4) Organizing Terms; (5) The Hierarchical Relationship; (6) The Associative Relationship; (7) Characteristics of Descriptors; (8) The Entry Vocabulary; (9) Scope Notes and Identifiers; (10) Thesaurus Format and Display; (11) Growth and Updating; (12) Computer Use; (13) Vocabulary Factors Affecting the Performance of Information Systems; and (14) Natural Language Systems. In several places, the text suggests topics upon which the instructor might elaborate, and it is noted that the guide could be used for a self-study program. A 10-item bibliography is included. (KM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PGI-85/WS/11
Paris, 1985

ED273297

THESAURUS CONSTRUCTION AND USE

A CONDENSED COURSE

PREPARED BY
F.W. LANCASTER

IR051622

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
C.Coudert-Schklowski
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

GENERAL INFORMATION PROGRAMME AND UNISIST
UNESCO

PGI-85/WS/11
Paris, March 1985

THESAURUS CONSTRUCTION AND USE
A CONDENSED COURSE

PREPARED BY
F.W. LANCASTER

GENERAL INFORMATION PROGRAMME AND UNISIST
UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

Reproduced by the Indian National Scientific Documentation Centre (INSDOC), New Delhi, for Unesco

Recommended catalogue entry :

Thesaurus construction and use / prepared by F.W. Lancaster /for the7 General Information Programme and UNISIST.- Paris : Unesco, 1985.- ii, 89 p. ; 30 cm.- (PGI-85/WS/11)

I Thesaurus construction and use
II Unesco. General Information Programme and UNISIST

© Unesco, 1985

CONTENTS

INTRODUCTION 1

UNIT . . . . .

1. PURPOSE OF VOCABULARY CONTROL 3

2. MAJOR COMPONENTS OF A CONTROLLED VOCABULARY 7

3. GATHERING TERMS 10

4. ORGANIZING TERMS 13

5. THE HIERARCHICAL RELATIONSHIP 15

6. THE ASSOCIATIVE RELATIONSHIP 17

7. CHARACTERISTICS OF DESCRIPTORS 18

8. THE ENTRY VOCABULARY 20

9. SCOPE NOTES AND IDENTIFIERS 22

10. THESAURUS FORMAT AND DISPLAY 23

11. GROWTH AND UPDATING 27

12. COMPUTER USE 29

13. VOCABULARY FACTORS AFFECTING THE PERFORMANCE OF INFORMATION SYSTEMS 33

14. NATURAL LANGUAGE SYSTEMS 38

EXHIBITS 40

BIBLIOGRAPHY 88

INTRODUCTION

The origin of this course was the Seminario Regional sobre Lenguajes de Indización organized under Unesco auspices by the Centro Argentino de Informacion Cientifica y Tecnologica (CAICYT) and held in Buenos Aires in August 1978. The author of the present volume was principal instructor at this seminar.

CAICYT later prepared a course, derived from the seminar itself, in the form of illustrations (actually transparencies) and accompanying text. This was published in 1981 (Curso sobre Lenguajes de Indización [y] Construcción de Tesauros. Unesco publication PGI.81/WS/31).

In 1984 Unesco entered into a contract with the author to prepare an English version of the course. This volume is the result.

The volume contains two principal components: a set of 84 "exhibits", which could be converted into slides or transparencies, and a text to explain and amplify the exhibits. It differs considerably from the Spanish version. In particular, it concentrates almost exclusively on the construction and characteristics of thesauri, whereas the seminar in Argentina branched out into other areas.

The text is divided into 14 topical units but no attempt has been made to make these "symmetrical": some are rather brief, others much longer.

The course is not intended as a programme of self-study, although it could be used for this purpose. Rather, it was prepared for the use of an instructor--the exhibits used as visual aids and the text as a guide to what might be included in lectures. The course should not be regarded as complete in itself. There are many places in which the text suggests topics that might be elaborated on by an instructor (e.g., differences between thesauri and lists of subject headings). It is expected that instructors will add further exhibits of their own to those supplied here, especially exhibits illustrating thesauri of special interest to the students attending a particular course. Moreover, it is important that many examples of thesauri of various types should be available for students to handle and examine. Also, it is highly desirable that each student should be given a copy of Unesco's Guidelines for the Establishment and Development of Monolingual Thesauri (Second edition, 1981).

Finally, a course on thesaurus construction must have a practical component. Each student--or small group--should be required to compile a small thesaurus--about 200 terms--in some subject area. The thesaurus must be structurally complete in the sense that it should include all types of components--BT, NT, RT, scope notes, etc., and all relationships should properly reciprocate. Moreover, it should observe the requirements of the Guidelines in all respects. To assist students in the task of building these thesauri, it would be necessary to supply a variety of sources--reference books, abstracts, etc.--from which they can draw their terms.

This volume is not a textbook. To keep it relatively short, some topics are dealt with only briefly. Greater detail on all of the topics covered, plus several related topics, can be found in Lancaster, Vocabulary Control for Information Retrieval (Second edition. Arlington, Va., Information Resources Press, 1985), which is presented in a sequence similar to that used in the present course.

In most cases, the instructions and examples given in this course conform to the Unesco Guidelines. Nevertheless, the course is not intended solely to illustrate use of these guidelines but to cover the broader subjects of construction and use of thesauri. Consequently, other widely used terminology, procedures, and practices, which may not fully conform to the Guidelines, are sometimes referred to and illustrated. These deviations should promote class discussion on the merits and disadvantages of the various approaches and practices in different contexts.

The author would like to thank the British Standards Institution for permission to reproduce pages from the ROOT Thesaurus.

UNIT 1. PURPOSE OF VOCABULARY CONTROL

Objectives

1. Illustrate the role the vocabulary plays in information retrieval systems.

2. Demonstrate some of the problems that occur when this vocabulary is not controlled.

3. Illustrate the major functions of a controlled vocabulary.

Exhibit 1

The term information retrieval is generally used to refer to the activities involved in locating documents (books, periodicals, articles, reports, and other forms) dealing with particular subject matter, and an information retrieval system consists of a group of activities and components designed to facilitate access to the subject matter of documents.

The major components are illustrated here. From the universe of available documents, those judged relevant to the interests of a particular community are selected and acquired by some information centre (which may be a conventional library).

Descriptive cataloguing and subject cataloguing or subject indexing are applied to the items selected. Conceptually, subject cataloguing is the same as subject indexing but the former term usually refers to the description of the subject matter of complete publications (e.g., books, periodicals) while the latter is more likely to apply to parts of publications (e.g., periodical articles). In this course, the term indexing refers to subject indexing although the principles discussed apply equally to subject cataloguing.

The first step in indexing, conceptual analysis, involves deciding what a document is about. The second step, translation, entails the selection of terms to represent what the document is about. These may be referred to as index terms or, since they describe the subject matter of documents, descriptors.*

The complete set of index terms used in a retrieval system may be referred to as the vocabulary or index language of that system.

At the heart of the retrieval system are two data bases, one consisting of the documents themselves, the other of entries representing these documents and providing access to them under the index terms selected to represent their subject matter. This second data base can be considered an index to the contents of the first.

In the case of most libraries, the data base of documents will consist primarily of books arranged on shelves while the index is likely to comprise entries within a card catalogue. In other situations, the index will be in machine-readable form (on magnetic tape or disk), on microfilm, or in printed book form, although other possibilities also exist.

---

* Guidelines pertaining to subject indexing may be found in the international standard ISO DIS 5963, Methods for examining documents, determining their subjects and selecting indexing terms.

In the exhibit, requests refers to information needs, arising among the population of system users, that are expressed as demands upon the system. The requestors will usually be seeking documents, or parts of documents, that discuss a particular subject, although they may, instead, be seeking the answer to a factual question.

In the information centre environment, what happens to incoming requests is similar to what happens to incoming documents. Conceptual analysis here refers to the interpretation of what a requester is really looking for, while translation is the representation of the requester's need in terms selected from the vocabulary of the system. The resulting representation of the information need can be considered a search strategy.

When a search strategy is applied to the index, those entries that satisfy the requirements of the strategy are "retrieved". This may later lead to the retrieval of the corresponding documents from the document store.

It should be obvious from Exhibit 1 that the vocabulary used in a retrieval system is a component of critical importance since the subject matter of both requests and documents must be represented by terms from this language.

In most retrieval systems the vocabulary used will be a controlled vocabulary. This need be nothing more than a limited set of terms that must be used by indexers and searchers. Put differently, the vocabulary prescribes which terms may be used and which not. In addition, however, controlled vocabularies will usually have some "structure" (i.e., the terms will be organized to show important interrelationships) but this is not essential to the definition.

Within the information retrieval context, the opposite of controlled vocabulary is natural language. That is, subject matter may be described by any words or phrases, without limitation, such as those occurring in documents themselves. It is possible to operate a retrieval system without any vocabulary control, and this will be discussed later in the course. However, certain problems do arise when no control is imposed on the vocabulary, as illustrated in Exhibit 2.

Exhibit 2

The terms here can all be considered to relate in some way to the failure of metals. Some refer to types of failure (e.g., fracture, rupture), some to possible causes of failure (e.g., corrosion, stress), some to properties of metals that may affect their tendency to fail (e.g., ductility, strength) and so on.

The terms are presented alphabetically as they might appear in a subject catalog or index, where they would be interspersed with many others on quite different topics.

For the person looking for information relating to the failure of metals, various problems are illustrated in this example:

1. Some terms listed are synonymous or nearly synonymous: "flaws" and "defects" for example.

2. Some words are ambiguous if taken out of context. For example, "cracking" can refer to the process of fracture or to a procedure used in chemical engineering (as in catalytic cracking) while such terms as "stress" and "fatigue" can apply to humans as well as to materials.

3. Terms whose meanings are related are widely separated alphabetically.

While it may be unlikely that anyone would want to perform a search on every aspect of failure, it is possible that someone would want to retrieve all information dealing with one form of failure, say fracture. The alphabetical sequence, unfortunately, separates many terms bearing on this topic: brittle fracture, cracking, ductile fracture, ductility, flaws, fracture, hydrogen embrittlement, and so on. Unless these terms are grouped or linked in some way, it may be very difficult for a searcher to think of all of them.

Exhibit 3

It is exactly these types of problems that the controlled vocabulary attempts to solve:

(a) Words or phrases that are synonymous or nearly synonymous are "merged": one is selected for use and some type of pointer from the other is inserted into the vocabulary.

(b) Words that have several possible meanings or contexts may appear more than once; parenthetical qualifiers or other devices are used to remove ambiguity.

(c) Terms whose meanings are related are linked together in some way. Links among some of the terms of Exhibit 2 are shown here. Two kinds of relationships are depicted. First, the permanent relationship between a thing and its kinds (i.e., genus and species): fracture and rupture are always kinds of failures. Second, a more transient relationship; for example, the process of failure may be induced by other processes, such as corrosion or fatigue, and it may be influenced by such materials properties as strength and brittleness. The permanent generic relationship is sometimes referred to as a paradigmatic or a priori relationship, whereas the more transient relationship (failure may be due to corrosion but is not always due to corrosion and corrosion does not always lead to failure) is sometimes referred to as a syntagmatic or a posteriori relationship.

Exhibit 4

The objectives of vocabulary control within an information retrieval system can be summarized as follows:

1. To promote consistent representation of subject matter by indexers and searchers, thereby avoiding the dispersion of related materials, through control (merging) of synonymous and nearly synonymous expressions and by distinguishing among homographs.

2. To facilitate the conduct of a comprehensive search on some topic by linking together terms whose meanings are related paradigmatically or syntagmatically.

Indexing tends to be more consistent when the vocabulary used is controlled. Indexers are more likely to agree on terms needed to describe a topic if they are selected from a pre-established list. The same is true of searching: it is easier to identify the terms appropriate to some information need if these terms must be selected from a definitive list. The controlled vocabulary, then, tends to bring the language of indexers and of searchers into coincidence.

A controlled vocabulary need be nothing more than a limited set of terms that must be used by indexers and searchers. More commonly, however, some structure will be introduced so that terms whose meanings are related are brought together or linked in some way. This helps both indexer and searcher to select the terms most appropriate to represent a particular topic and to identify all the terms needed to perform a fairly comprehensive search on the subject.

The controlled vocabulary--of which the thesaurus is one type--is most often used to standardize descriptors or subject headings representing the contents of publications or the interests of individuals (i.e., search strategies and interest profiles used in Selective Dissemination of Information (SDI)). In principle, however, such a tool can be applied in any situation in which the standardization of terminology is needed.

UNIT 2. MAJOR COMPONENTS OF A CONTROLLED VOCABULARY

Objectives

1. Show that a controlled vocabulary must have both classified and alphabetical components.
2. Show a number of ways in which these components can be presented.
3. Illustrate a conventional thesaurus structure.
4. Demonstrate that the thesaurus embodies a complete hierarchical classification.

Exhibit 5

A controlled vocabulary should have two complementary components: a systematic (classified) presentation of terms and an alphabetical presentation.

Exhibit 6

This shows a partial classification of terms describing types of libraries. Here it is presented in the form of a family tree, as used in genealogy, so that it is easy to see that junior college libraries is a subdivision of college libraries, which is a subdivision of academic libraries, and so on. Subdivision is on a strict genus/species basis. Junior college libraries are kinds of college libraries which are kinds of academic libraries.

Exhibit 7

The family tree display works well with relatively small hierarchies of terms but very large hierarchies are difficult to display in this way because it may not be possible to present the entire hierarchy on a single page. Moreover, the family tree display wastes space.

In this exhibit, an identical hierarchical classification is presented. Here, however, the terms appear in an arrangement that resembles a conventional bibliographic classification scheme, as used in libraries (such as the Universal Decimal Classification or that of the Library of Congress). Various levels of indentation are used to indicate the levels of the hierarchy.

Exhibit 8

This looks even more like a bibliographic classification scheme because each term has now been given a classification number. These numbers are referred to as the notation of the classification. Some students of library science are confused in that they believe a notation is essential to a classification scheme. It is not. Exhibit 8 is no more a classification scheme than Exhibit 7 or, indeed, Exhibit 6. Exhibits 6 and 7, as far as they go, are true classification schemes covering a part of the terminology of "libraries". The notation serves merely to keep the scheme in sequence and to provide a convenient, shorthand symbol for each term--useful, for example, in organizing books on library shelves.

BEST COPY AVAILABLE

Exhibit 9

Exhibits 6-8 present terms in a systematic (classified) manner but do not give the alphabetical approach. They must be supplemented by some type of alphabetical index to show a user where in the classification scheme a particular term appears. This exhibit is a simple example of alphabetical index entries based on the classification of Exhibit 8. If the classification is presented without notation (Exhibits 6 and 7), of course, the most the index can do is to refer to the page on which a particular hierarchy appears--which illustrates one of the disadvantages of not having a notation in a classification scheme.

Exhibit 10

In Exhibits 6-8, the principal arrangement of terms is classified and it is overtly classified. The alphabetical index is really a supplement, although it is a supplement essential to the effective use of a classification scheme of any size. A classification scheme has two separate but interrelated components, one classified and one alphabetic.

Exhibit 10 is in alternative display of the same set of library terms. It is overtly alphabetical. However, the hierarchical classification that is overt in Exhibits 6-8 has been built into the alphabetical organization covertly by means of cross-references. Here, a genus is referred to as a broader term (BT) and a species as a narrower term (NT). Thus, we can see that ACADEMIC LIBRARIES has two species and one genus, while LIBRARIES has four species and no genus (i.e., it is the term at the top of the hierarchical tree).

Exhibit 10 is a rather simplified illustration of how terms are presented in an information retrieval thesaurus. The thesaurus differs from the classification schemes depicted earlier in that:

(1) The arrangement is overtly alphabetical and covertly classified.

(2) Alphabetical and classified components are presented in a single sequence.

(3) At each entry point, only one hierarchical level upwards and one downwards is given. Thus, one can see from the ACADEMIC LIBRARIES term that COLLEGE LIBRARIES is one level down in the hierarchy but one must go to COLLEGE LIBRARIES to see that JUNIOR COLLEGE LIBRARIES is a lower level in the same hierarchy. This "one step up, one step down" feature is not essential in thesauri but it is one that conserves space when the tool is presented on paper.

(4) The BT/NT relationship is reciprocated symmetrically. If term B is shown to have an NT, term Y, then term Y must show B as its BT.

While covert, the classification embedded in Exhibit 10 is just as complete and correct as the classification revealed more openly in Exhibits 6-8. Indeed, Exhibit 10 is derived directly from Exhibit 7.

Moreover, since Exhibits 6, 7, 8 and 10 present an identical classification scheme, one display could be derived from another automatically. That is, a computer program could be written to derive 10 from 6, 7, or (most easily because of the notation) 8, and a computer program could be written to derive 6 or 7 from 10. Exhibit 8 could also be derived automatically from 10. In this case, however, the program would need to be extended to allow the notation to be arrived at algorithmically.

Note that the thesaurus as used in information retrieval is quite different from the literary thesaurus as exemplified by that of Roget. It is also different from the list of subject headings as traditionally used in libraries. The list of subject headings does not embody a complete classification scheme. Furthermore, subject headings are descriptors more "complex" than the descriptors of a thesaurus in that they tend to be more "pre-coordinate" (see discussion of Exhibits 73-75).

Exhibit 11

In all of the "libraries" examples used so far, only one type of relationship has been shown: that among terms appearing in the same hierarchy. But terms from different hierarchies can also be related, as illustrated by the pairs linked with arrows in 11(a).

A controlled vocabulary will also link terms that are closely related, although appearing in different hierarchies. In the thesaurus, this is achieved by means of the related term reference (RT) as shown in 11(b). The BT/NT references link terms from the same hierarchy while the RT reference links terms drawn from different hierarchies. This is discussed in more detail later.

UNIT 3. GATHERING TERMS

Objectives

1. Distinguish between "top down" and "bottom up" approaches to thesaurus construction.

2. Define "literary warrant" and "user warrant" and show the importance of these principles.

3. Show how these principles can be implemented in gathering terms for the construction of a thesaurus.

Exhibit 12

Before building a completely new thesaurus, an information centre would do well to review existing thesauri to see if one will meet local needs. In some cases, an existing thesaurus can be modified to make it usable. This may mean expanding it in various subject areas. A special form of expansion is the building of a micro-thesaurus that fits within the structure of a more general tool (e.g., a thesaurus on diabetes fitting within the hierarchical structure of an existing medical thesaurus).

In other situations, a suitable thesaurus may exist in another language and can be translated into the language of the information centre.

Building a thesaurus anew can be a costly undertaking. It should only be done when one is certain that no existing tool can be adopted or adapted.

Exhibit 13

In the building of a thesaurus, four major steps can be identified. The first involves gathering the raw material--the terminology of the field to be covered.

Exhibit 14

Two approaches to thesaurus construction are shown. In the top-down or "deductive" approach, which frequently involves use of committees, the major categories of terms are identified and subdivided from the top down. Thus, a group of librarians, building a thesaurus on library science, recognizes the need for a category "types of library". Using "libraries" as top term, they begin to subdivide the category along the lines shown in Exhibit 6.

Major problems exist with this approach:

1. It is difficult to think of all the categories or hierarchies that will be needed.

2. The characteristic used to subdivide a genus may not lead to a classification that best meets needs of users. For example, the class "toys" could be subdivided by material, by colour, by age of intended audience, by method of locomotion, and so on. For a toy manufacturer, one of these characteristics (perhaps age) may be much more important than some of the others.

BEST COPY AVAILABLE

The empirical (i.e., bottom-up or "inductive") approach to thesaurus construction is the opposite extreme and it tends to be much more reliable. Terms occurring in the field are collected from various sources and a category or hierarchy of terms is formed only if it appears to be important or useful. For example, a subdivision of toys by colour (blue toys, red toys), while theoretically possible, may have no significance to the toy manufacturer for whom the thesaurus is created.

In the empirical approach, two important principles--literary warrant and user warrant--can be identified.

Exhibit 15

The principle of literary warrant (also known as bibliographic warrant) was enunciated by Wyndham Hulme more than seventy years ago. The context was that of bibliographic classification: a term (class number) is justified only if literature on this topic is known to exist.

Applied to thesaurus construction, one can extend the principle: a term is justified only if it occurs within the literature of a subject with some degree of frequency.

The most efficient way to collect the terminology of a field is to go to those reference sources likely to contain a high concentration of terms. The sources must be reasonably up-to-date.

Perhaps the best source of all will be an abstracting publication devoted to the field or the appropriate section of one of broader scope (e.g., the educational psychology section of Psychological Abstracts).

Working in a purely manual mode, the thesaurus compiler would go through the abstracts, recording words and phrases that seem important in describing the subject matter of the field. These could be underlined in the abstracts or transcribed onto cards. Eventually they should be put into machine-readable form to allow manipulation by computer.

It makes most sense to work backwards in time, starting with the latest issue of the abstracts. The compiler stops collecting terms when a point of diminishing returns seems to have been reached; i.e., when very few new terms are encountered.

Collecting terms manually in this way is tedious and time-consuming. The process may be expedited considerably if the text of the abstracts can be obtained as a machine-readable data base. In this case, programs can be written to print a list of words occurring in titles and abstracts in descending order of frequency. The same can be done for word pairs or triples occurring significantly often in the data base. Moreover, it would also be possible to develop printouts of word co-occurrence in abstracts, which might be valuable in selecting RT's for use in the thesaurus.

Exhibit 16

Selecting terms from the literature (literary warrant) is important. But it is also important that the terms collected actually represent the needs of the users of the information service that is to apply the thesaurus. Certain terms may occur in the literature but be of little interest to a particular group of users. For example, the users may need less specificity than the literature indicates for certain types of terms.

The thesaurus compiler would do well to identify terms indicative of the subject interests of users, as well as drawing terms from the literature. These are complementary rather than alternative approaches.

If the information service has existed for some time and has collected records of literature searches performed for users in the past, these are likely to be a rich source of terminology. Questionnaires can also be used (sent to all prospective users or to a random sample, depending on numbers involved) to gather narrative statements indicative of the current interests of the people to be served. In the case of a research community, titles/abstracts of user publications constitute a valuable source of terms.

Under certain conditions, it may be possible to combine the literary warrant and the user warrant principles in a single operation. Representative users are supplied with copies of publications that closely reflect their current professional interests. The users then underline the terms that make these publications of interest to them.

UNIT 4. ORGANIZING TERMS

Objectives

1. Introduce the standards for thesaurus construction.
2. Show how terms can be organized into categories and into hierarchies within categories.
3. Demonstrate how an alphabetical thesaurus display can be derived systematically from the hierarchical display.
4. Show alternative possibilities for displaying the alphabetical component of a thesaurus.

Exhibit 17

The first thesaurus for information retrieval dates back to 1959. Much experience in thesaurus construction was gained in the early 1960's, especially in the United States. Gradually this experience was recorded, leading to the publication of guidelines and later to national standards for thesaurus construction (e.g., in the United States, the United Kingdom, France and Germany).

Unesco issued guidelines for the construction of monolingual thesauri in 1970 and these led to the international standard ISO 2788, published in 1974. A revised edition of the Unesco guidelines, issued in 1981, formed the basis for a draft of a revised edition of ISO 2788.

Unesco guidelines for multilingual thesauri appeared in 1976 and were revised in 1979, leading to ISO DIS 5964 (1983).

The thesaurus features discussed in this course, unless otherwise noted, are compatible with the Unesco Guidelines for the Establishment and Development of Monolingual Thesauri (Second edition, 1981).

Exhibit 18

Once terms are collected--from the literature, from users, or both--they need to be organized into major categories and into hierarchies within these categories. Useful inter-hierarchical relationships must be established. Finally, the systematic structure must be "inverted" to form an alphabetical arrangement.

Exhibit 19

Suppose one were constructing a thesaurus on "irrigation" and had collected on cards a large assortment of terms. The cards can now be put into piles, so that all cards in a particular pile are of the same "type" (i.e., fall in the same category). Presumably, one category will be for types of irrigation, one for types of crops to be irrigated, one for soil types, one for climatic zones, and so on.

Exhibit 20

Once the categories have been identified, the next step is to organize each into hierarchies. In most cases, the entire category can usefully be organized into one

large hierarchy, as might be the case for crops, which are divided into major types, major subtypes, and so on. In a few cases, however, it may make more sense not to attempt a single hierarchy. For example, one of the categories first identified may be "properties". However, the compiler later decides that this general term is too broad and not really useful as the "top term" in a hierarchy. Instead, he constructs several separate hierarchies: physical properties, mechanical properties, shape, colour, and so on.

This exhibit shows a partial development of the "crops" hierarchy. As far as it goes, it is a true hierarchical classification based on the genus/species relationship: oranges are kinds of citrus fruits, citrus fruits are kinds of fruits, fruits are kinds of crops.

Exhibit 21

The complete set of hierarchies developed from the irrigation terms would constitute a classification of the terminology of irrigation. When the hierarchies are established to the satisfaction of the compiler, the classification is inverted to create the alphabetical thesaurus. Each term becomes an entry and its hierarchical relationships are retained, and expressed as BT's and NT's.

This exhibit shows a group of thesaurus entries derived from Exhibit 20. All of the genus/species relationships have been retained and they fully reciprocate: because "lemons" shows "citrus fruits" as its BT, "citrus fruits" must show "lemons" as one of its NT's. Exhibits 20 and 21 are merely two different ways of displaying the same classification scheme: 21 was derived from 20 and 20 could be derived from 21.

Exhibit 22

The one step up, one step down display (as described earlier) is used in Exhibit 21. This is recommended, if for no other reason than to save space in a printed thesaurus.

Other forms of hierarchical display are sometimes used, as shown in Exhibit 22. On the left, every level of narrower term is given, in one alphabetical sequence. This has nothing to recommend it: it wastes space and does not distinguish the hierarchical levels. On the right is a kind of "alphabetico-classed" display, which has the advantage of showing all hierarchical levels at a glance. The only disadvantage is that it will lead to a very bulky printed tool. For a relatively small and specialized thesaurus, however, it may be the best form of display.

Exhibit 23

If adopted, the alphabetico-classed type of display should probably be explicit in identifying hierarchical levels: NT1 is a narrower term at first level, NT2 at second level, and so on.

UNIT 5. THE HIERARCHICAL RELATIONSHIP

Objectives

1. Distinguish the "generic", "partitive" and "instance" relationships.
2. Demonstrate how the BT/NT relationship is to be used.
3. Show how the BT/NT relationship may be displayed in a printed thesaurus.

Exhibit 24

The BT/NT relationship, under almost all circumstances, will be a genus/species relationship (i.e., generic). To be a legitimate NT, a term must represent a kind of the thing identified by its BT.

Exhibit 25

The whole/part (partitive) relationship, usually, will not be treated as BT/NT. For example, in a thesaurus on furniture construction, the term "table legs" should appear as an NT under "legs", not under "tables".

By convention, however, certain partitive relationships are frequently treated as though they were generic. Parts of the body and geographical names are the prime examples. It makes sense to make "aorta" an NT under "heart" even though it is a part rather than a kind of heart, and likewise for the geographic situation. The Unesco Guidelines explicitly recognize two other situations in which the partitive relationship should be treated as generic: fields of study (Physics can be an NT of Science) and social structures (Divisions can be an NT under Armies).

Exhibit 26

The Unesco Guidelines allow a thesaurus to recognize two kinds of BT/NT relationships, one generic and one partitive, so that "part" terms can be grouped with the "whole" term to which they relate. In this case, a distinction is made between NTG (narrower term generic) and NTP (narrower term partitive), and the two groups of terms are listed separately. The relationship reciprocates: in this example, "warships" will have a BTG "ships" and "hulls" will have a BTP "ships". This procedur is useful in that it avoids the introduction of rather artificial terms to serve as a BT for the parts terms; e.g., "ship parts" or "ship components".

Exhibit 27

The Guidelines also allow the "instance" relationship to be treated as BT/NT where this is useful. An "instance" NT will be some type of proper name--of a person, group, building, and so on. "Coventry Cathedral" is not really a kind of cathedral (a "kind of" would be, for example, Gothic) but an example (instance).

Exhibit 28

In a properly constructed faceted classification scheme, the principle of division (or "characteristic") used to derive the facets should be clearly and explicitly stated.

Exhibit 29

In a thesaurus, however, such distinctions are not always necessary. This entry is perfectly acceptable even though the NT's represent three different principles of division (i.e., facets): application, country and style.

Exhibit 30

Nevertheless, the explicit identification of facets in a thesaurus has much to recommend it, particularly where many terms are involved. The type of entry illustrated would avoid a long list of heterogeneous NT's under the term ARCHITECTURE.

There are other possible ways of handling the situation, e.g., by introducing the terms ARCHITECTURAL STYLES (NT Baroque architecture, etc.) and ARCHITECTURAL APPLICATIONS (NT Domestic architecture, etc.).

Exhibit 31

Within a thesaurus, a term will normally have only one BT (i.e., it will appear in one hierarchy only). In some cases, however, it may legitimately be given two or more BT's. The more general the thesaurus, the more likely this is to occur. Given a choice, it is best to put a term in the hierarchy in which it always belongs: a diamond can be a cutting tool, but it is always a precious stone.

Nevertheless, the hierarchical relationships should reflect the interests of the community to be served. For example, in a thesaurus for the metal industry there would be little need for a precious stones hierarchy. Here, presumably, "diamonds" would appear only under "cutting tools".

Exhibit 32

Some thesauri will show, for each term, not only its BT but also its TT (i.e., the "top term" in the hierarchy in which it appears). In this example (see Exhibit 20), the BT of "oranges" is "citrus fruit", while its TT is "crops".

UNIT 6. THE ASSOCIATIVE RELATIONSHIP

Objectives

1. Distinguish between the hierarchical (BT/NT) and the associative (RT) relationship.

2. Give examples of the types of terms that should perhaps be linked by RT's.

Exhibit 33

Precise rules can be formulated to govern the BT/NT relationship. The RT, or "associative", relationship is much less susceptible to regulation. The only thing that can be said fairly definitely is that any two terms that can be considered related in meaning, within the context of a particular thesaurus, can be linked by the RT reference, providing that they do not appear in the same hierarchy. That is, the RT is an inter-hierarchical connector rather than an intra-hierarchical one. It quite unnecessary to use an RT to link domestic architecture with ecclesiastical architecture or naval architecture. These are siblings--co-equal members of the same hierarchy--and the relationship is made clear by moving up to the broader term, in this case architecture.

On the other hand, the arrows in this exhibit show terms that might legitimately be linked by RT to ecclesiastical architecture since these should be from quite different hierarchies.

The RT should link two hierarchies at the most appropriate levels of these hierarchies. Suppose, for example, that churches has several NT's: cathedrals, abbeys, and so on. It would be wrong to link ecclesiastical architecture to these NT's; since the reference applies to churches in general, it should be made at this broader level.

Unlike the BT/NT relationship, it is not essential that the RT relationship should reciprocate. One is not obliged to make ecclesiastical architecture an RT under iconography just because the reverse reference occurs. Nevertheless, reciprocation will usually be desirable.

Exhibit 34

Some relationships that would frequently be expressed through the RT are illustrated. The list is not complete although it perhaps includes the most importa and frequent relationships. In all cases illustrated, the reciprocal would also apply (e.g., cause/result as well as result/cause).

UNIT 7. CHARACTERISTICS OF DESCRIPTORS

Objectives

1. Define a legitimate descriptor in terms of:

   (a) word form

   (b) singular/plural

   (c) word sequence.

2 Give guidance on "compounding", i.e., descriptor length and complexity.

Exhibit 35

Descriptors will almost always be nouns or noun phrases. Most noun phrases will be adjectival phrases but some prepositional phrases may also occur. Verbs in the form of infinitives or participles should not be used but verbal nouns (e.g., cutting) can appear as descriptors. Unattached adjectives are best avoided. Descriptors should always be in direct rather than inverted form. This misleads some students, who would like to group related terms alphabetically--e.g., all book terms together. This is quite unnecessary--grouping is achieved in other ways (e.g., moving up to the genus "books").

Exhibit 36

Most nouns are quantitative or "count" nouns. They represent entities measured in numbers ("how many?") and should always be in the plural form. The less common nouns of volume represent items measured in amount ("how much?") and should appear in the singular.

An exception can be made for nouns usually thought of as singular or thought of as singular in a particular context. It makes sense to say "eyes" and "ears" but to use "nose". In a thesaurus on ships one might use "bridge" rather than "bridges" on the grounds that there is usually only one bridge per ship.

Exhibit 37

Avoid needless redundancy within a specialized thesaurus. In a thesaurus on library science, the term "cooperation" can reasonably be interpreted as meaning "library cooperation", "buildings" as meaning "library buildings", and so on. It is not really useful to find half the descriptors in the thesaurus beginning with the word "library".

Exhibit 38

One of the most difficult problems facing the thesaurus compiler is deciding to what extent to split a phrase into its components. One extreme is to keep "automobile engine noise" as a single term; the opposite extreme would be to break it down into single words. Something in between is also possible. The choice will depend on several factors, including the scope of the thesaurus and the conventions of the language involved (e.g., German versus English practice).

BEST COPY AVAILABLE

Exhibit 39

The Unesco Guidelines give some rules to assist these decisions, based on the principle of "syntactical factoring". These are fairly complex and should be studied within the Guidelines themselves. This exhibit attempts a more simplified approach. It does not cover all cases but it perhaps deals with the more important.

In general, phrases whose components appear in their own right in the thesaurus should usually be split: "traffic noise" is an example if both "traffic" and "noise" are also needed. The specific idea of "traffic noise" can be expressed by indexing an item under both component terms.

Phrases that represent two different principles of division (i.e., "characteristics") should always be split. "Children's picture books" embodies two different characteristics: audience intended and physical format.

Terms that appear to represent a species but really express another relationship should always be split. "Gas turbine corrosion" is not a species (kind) of corrosion (unlike, say, stress corrosion)--the gas turbine is the location of the corrosion only.

Compounds that generally should not be split:

- Those in which a word has only one context in a particular thesaurus--in a thesaurus of library science, "acidity" may exist only in connection with paper.
- Those in which the meaning of a word in the compound is different from its meaning alone: tiger lilies have nothing directly to do with tigers.
- Those in which the components could be ambiguous: "plates" and "glass" could represent glass plates or plate glass.
- Those that masquerade as members of a well-known genus but mean something quite different: clothes horses are not real horses.
- Those that incorporate a proper name.
- Those occurring so frequently in a particular field that splitting of the compound, while possible, would cause irritation to users, as in the case of "pressure vessels" within mechanical engineering.

UNIT 8. THE ENTRY VOCABULARY

Objectives

1. Demonstrate use of the use reference.
2. Show how this reference may be used to control synonyms and near-synonyms.
3. Show how this reference may be used to refer from a specific term, not used in indexing, to a more general one.

Exhibit 40

A thesaurus will include references from certain terms not used for indexing or searching (nondescriptors) to the terms used in their place (descriptors). The non-descriptors point to the descriptors. They provide additional access or entry points to the vocabulary. Collectively they can be referred to as an entry vocabulary. As shown in the exhibit, there are two major types of entry terms.

Exhibit 41

Exact synonyms should be controlled through the use reference. One form of synonym is an abbreviation or acronym. Another type is that reflecting different national conventions (e.g., American versus British usage). Use as a descriptor the term most likely to be thought of by the audience to be served. For a British audience, railways is to be preferred over railroads; the general public may be more likely to look under stamp collecting than under philately.

Try to be consistent. It is confusing if abbreviations or acronyms are used in some cases but not others. However, an acronym should always be used when the full form is little known, as in the case of lasers.

Near-synonyms occur more frequently than true synonyms. A term, A, can be considered "synonymous" with another, B, when the distinction between them is unclear or not worth making in a particular context. "Sound" is not exactly synonymous with "acoustics" but these terms could be treated as synonymous in certain thesauri.

It is not necessary to refer from inverted forms to direct forms because the additional access point will usually be provided in other ways. For example, the entry Books, picture use Picture books is redundant since the term "Books" will list the descriptors for all types of books as its NT's. A possible exception can be made in the case of prepositional phrases (e.g., Equations of motion).

Terms representing opposite extremes on a continuum of values are usually treated as though synonymous. For example, an article discussing the elasticity of some material can be considered to deal equally with its inelasticity. These pairs are sometimes referred to as "quasi-synonyms".

Mapping can be one-to-many as well as one-to-one. In example 7, the topic "domestic architecture" is to be indexed and looked for under both "architecture" and "houses".

Terms considered too specific to be used for indexing purposes can be mapped to the appropriate more general term. For example, there is literary warrant for "white noise" but a particular thesaurus may not require this degree of specificity.

BEST COPY AVAILABLE

Exhibit 42

The "use" reference must reciprocate. If the reference A use B exists, the reciprocal B UF A, where UF represents "used for", must also exist. In the case of one-to-many mappings, some symbol is needed to indicate that a term is only one of two or more terms referred to. In this example, under Houses, the dagger after UF shows that the term "houses" is one of two or more terms referred to from "domestic architecture".

UNIT 9. SCOPE NOTES AND IDENTIFIERS

Objectives

1. Illustrate types of scope notes and show their purpose.

2. Show how homographs can be distinguished in a thesaurus.

3. Distinguish between a descriptor and an identifier.

Exhibit 43

When the meaning of a term may be unclear, it can be given a scope note (SN) in the thesaurus. This may be a true dictionary definition, as in the example of permafrost. Alternatively, the SN may merely indicate how a term is to be interpreted within a particular thesaurus and/or distinguish it from other terms with which it might be confused.

Exhibit 44

Some words may have more than one meaning, even within the context of a single thesaurus; i.e., they are homographs. It is frequently possible to distinguish different meanings by use of a parenthetical qualifier, without the need for a full scope note. Unlike the scope note, the parenthetical qualifier is an integral part of the descriptor. For example, the term "tanks" alone does not exist. It has been split into two separate terms: TANKS (CONTAINERS) and TANKS (WEAPONS). Of course, other methods of resolving such ambiguities exist and may be used in some thesauri.

Exhibit 45

The descriptors organized within a thesaurus tend to be nouns or noun phrases but may not include very many names. Nevertheless, in many subject fields, names of people, places, organizations or objects will also be needed to represent the subject matter of documents. Rather than greatly expanding the hierarchical structure of the thesaurus, such names may be "controlled" (i.e., standardized) in a separate vocabulary. The names treated in this way are frequently referred to as identifiers and the list used to control them will usually be just that--a list, with perhaps some cross references, but no real "structure". Examples of the types of terms treated as identifiers in various information services are given here.

This does not mean that all names are to be treated as identifiers in all thesauri. Clearly, a political science thesaurus will need names of countries, and groups of countries, as an integral part of the thesaurus structure. A thesaurus in education, on the other hand, has less need for geographic descriptors: names of countries can be handled through the identifier list. In fact, in a subject area in which such names will occasionally be needed it is perhaps best to adopt some existing authority for the names. In the United States, the obvious source would be the publications of the Board on Geographic Names.

For a thesaurus on, say, English literature, names of authors will obviously be needed. But it will be impractical to try to include all such names in the thesaurus structure. It will be enough to include names of the really major authors as descriptors, the others being treated as identifiers. In this case, a rather obscure nineteenth century essayist might be indexed under relevant descriptors--ESSAYISTS and NINETEENTH CENTURY--as well as under his name from the identifier list.

UNIT 10. THESAURUS FORMAT AND DISPLAY

Objectives

1. Illustrate the major components of a thesaurus entry.

2. Illustrate alternative ways in which a thesaurus can be displayed.

3. Demonstrate, with examples, graphic display and the thesaurofacet.

Exhibit 46

If the entry under a term has every customary element, it will look like this. The elements are presented in the sequence recommended by the standards. Within each list of terms referred to (NT, RT, etc.), terms are listed alphabetically. Most entries in a thesaurus will not have every element; in particular, scope notes will be the exception rather than the rule.

Exhibit 47

As described earlier (Unit 2), the alphabetical organization of terms in a thesaurus should embody a true hierarchical classification scheme. Nevertheless, many thesauri include some form of explicit classification scheme as a second display to complement the alphabetical display. Other displays are also possible.

The UNBIS Thesaurus used by the Dag Hammarskjöld Library, United Nations, New York, incorporates four complementary displays of terms. The alphabetical display, illustrated in this exhibit, is conventional, except for its inclusion of TT's (explained in Exhibit 32) and category numbers (e.g., 04.02.01, under IRRIGATION is the category in which this term appears).

Exhibit 48

The category numbers provide entry points to a categorized list of terms. This can be looked upon as a very broad classification scheme. Within each major category (e.g., Crops), however, the terms are listed alphabetically rather than hierarchically. As seen here, the display includes use references and scope notes.

Exhibit 49

The third display is a "hierarchical list". Here, each "top term" is arranged alphabetically and, below it, the entire hierarchy appears, using indentation to show the various levels.

Many thesauri will include some form of "systematic" display to complement the alphabetical display--something resembling Exhibit 48 or Exhibit 49. The inclusion of two different displays of this kind is less common.

Exhibit 50

Most thesauri, including the UNBIS Thesaurus, will include some form of permuted display of words: either KWIC (keyword in context) or KWOC (keyword out of context). Such an index shows, for any word, all of the occurrences of that word among the descriptors. It is a useful supplement because it reveals word occurrences (and

therefore "topic" occurrences) that might not be found easily in the alphabetical or other displays.

Exhibit 51

An alternative method of presenting a thesaurus is in the form of a "graphic display". The type illustrated, drawn from the Unesco Guidelines, is known as an "arrowgraph". In essence, it is a hierarchical display shown pictorially. The top term in the hierarchy, cameras, is the centre of the display and other terms radiate from it to show levels of the hierarchy. Thus, 35mm cameras is subordinate to miniature cameras which is subordinate to still cameras which is subordinate to cameras. Dotted lines link terms to other, related displays. A volume of such arrowgraphs somewhat resembles a road atlas, with one map leading to another. Proponents of the graphic display claim that, once you are used to the format, it is easier to survey a whole family of terms than with any other format.

Exhibit 52

The arrowgraphs, basically hierarchical, constitute the classified component of a thesaurus. Clearly, some alphabetical approach is also needed. This could be simply an alphabetical index. In this example, however, the alphabetical component is in fact in the form of a conventional thesaurus with one addition: the position of each term in the arrowgraphs--map number and quadrant reference--is given.

Exhibit 53

The Unesco-produced SPINES Thesaurus includes a completely different type of display that may also be considered "graphic". This thesaurus incorporates a number of unique features. The alphabetical component explicitly identifies all levels of BT and NT for each descriptor, as well as giving scope notes, RT's and use references. A distinction is made between the use and the see reference (whose reciprocal is sf, seen from), the latter giving a choice of descriptors to replace the nondescriptor (entry term). Inclusion of a see reference as well as the use reference is unconventional and is not recommended in the Unesco Guidelines. Students would be well advised to ignore this distinction.

Exhibit 54

The graphic display uses polygons, sub-polygons and sub-sub-polygons in an attempt to clarify various hierarchical levels. Related polygons are linked by an "associative relation," a type of RT reference. A complete page of the graphic display ("chart") may include several interrelated polygons, although the example given shows only one complete polygon.

Exhibit 55

As emphasized earlier, all thesauri will have a classified component--achieved by means of the BT/NT structure, a graphic display of hierarchy, or some other means.

The thesaurofacet is a type of controlled vocabulary that combines the alphabetical thesaurus with a faceted hierarchical classification, including notation.

Some entries from a hypothetical thesaurofacet are shown. Note that these two parts fully complement one another. The faceted component takes care of the hierarchical relationship (i.e., the BT/NT structure), while all other relationships appear in

BEST COPY AVAILABLE

the thesaurus component. The thesaurus component gives the notation for each term so that the user can go to the faceted component to see the hierarchical relationship. In the faceted classification, a term appears only once. However, if a term can legitimately appear in more than one hierarchy, the secondary relationships are given in the thesaurus, using the convention BT (A), which stands for "additional broader term." The thesaurofacet has an obvious advantage over other thesaurus forms: it can be used for arranging books on the shelves of a special library as well as indexing the items in a data base. Moreover, shelf arrangement and data base will be fully compatible.

Exhibit 56

The Unesco Thesaurus can be considered a type of thesaurofacet. Here is the faceted component with each facet explicitly identified ("by crimes against property", "by crimes against public order").

Exhibit 57

In this thesaurus, however, the alphabetical and faceted components are not completely complementary as shown in the hypothetical example (Exhibit 54). While the alphabetical component gives much information not in the faceted, it also repeats some of this information (e.g., BT, TT, UF).

Exhibit 58

The ROOT Thesaurus of the British Standards Institution is perhaps the most elaborate of the tools to combine a faceted classification with an alphabetical display of terms. The facets into which a subject is divided are clearly and explicitly indicated ("by property", "by application", and so on). The thesaurus employs some special conventions, as follows:

= indicates a non-preferred synonym

*< indicates an additional broader term from elsewhere in the display (no example in this particular exhibit)

*> indicates an additional narrower term from another part of the display (e.g., "drainpipes" appears in RMF.N as well as being considered a type of pipe under NJV/NJX; "rigid pipes", on the other hand, at NJV.B, does not appear elsewhere in the classification).

*- indicates an additional related term from another part of the display

Exhibit 59

The alphabetical section gives one level of hierarchy only. The following conventions are adopted:

= non-preferred synonym

< broader term in the same part of the display (e.g., "pipes" is shown to be a narrower term under "fluid equipment")

> narrower term in the same part of the display (e.g., "rigid pipes" is a narrower term under "pipes")

- related term in the same part of the display

*< broader term in another display

*> narrower term in another display ("drainpipes" was the example used previously)

*- related term in another display

Nondescriptors are handled in one of two ways, as in these examples:

Pipe threads

→ Threads NWC/NWD

indicates that the former topic is to be indexed under the latter term, while,

** Asbestos cement pipes NJW.S

→ Pipes

+ Asbestos cement

indicates a "synthesized term": the term "asbestos cement pipes" is not to be used; instead this concept is to be indexed under "pipes" and under "asbestos cement" (note that the entry under "pipes" shows that this term + "asbestos cement" is to be used to represent "asbestos cement pipes"). The special symbols used in place of BT, NT, RT and use make the structure of this thesaurus language-independent. However, use of such symbols is not advocated by the international standards.

UNIT 11. GROWTH & UPDATING

Objectives

1. Show how a simple test for "completeness" of a draft thesaurus may be performed.

2. Define the ways in which a thesaurus can be kept up to date.

Exhibit 60

A thesaurus will not be static; it will continue to grow as new terms are needed to accommodate topics not previously encountered. Usually this will mean developing existing hierarchies in more detail, thus making the vocabulary increasingly specific.

A hypothetical growth curve is shown. Assuming that the thesaurus is built by drawing terms from the literature, growth will at first be very rapid--virtually every new item (e.g., abstract) will generate several new terms. Eventually, however, it will begin to level off and a point will be reached where many new items must be examined before any new terms are encountered. At this point, the collecting of terms can be suspended: those collected should now be organized to produce a first draft of a thesaurus.

Exhibit 61

It would be desirable to have the draft reviewed, perhaps by subject specialists as well as the information specialists who are to use it. Before it is applied to actual operations, one might also want to test its adequacy.

A simple test is illustrated. Pick a random sample of, say, 100 recently published papers or reports of the type to be indexed into the data base. One (or more) of the information specialists indexes these in two steps. The indexer first jots down the results of his "conceptual analysis" as a series of words or phrases representing the topics that he feels should be covered in the indexing. This conceptual analysis is done for all 100 items in the sample. Next, the draft thesaurus is consulted and an attempt is made to translate the conceptual analysis into the controlled vocabulary.

Exhibit 62

A hypothetical example is shown in this exhibit. In this case, all of the concepts listed by the indexer can be translated into thesaurus terms, although a one-to-one correspondence is not always possible. For example, two "concepts" translate to one term or one concept to two terms. Only in one case is there a variation in specificity: the thesaurus includes "wastewater" but "sewage water" is really a special kind.

Exhibit 63

The results of the test on all sample items can be expressed statistically. In this hypothetical case, the results look good. Very few of the concepts are not translatable. Almost all can in fact be translated at the required level of specificity. A test of this kind may actually identify some precise deficiencies that can be corrected--perhaps more specific terms for types of water that may be used in

BEST COPY AVAILABLE

irrigation seem needed. Once this has been corrected, the thesaurus seems ready for application.

Exhibit 64

As it is applied, of course, the thesaurus will continue to grow. It must be kept up-to-date by the addition of new terms. The information service using the thesaurus must establish procedures whereby indexers can add new descriptors as needed, at least on a provisional basis. In the case of a large information service, some "vocabulary review group" may be formed to decide whether these "provisionals" should be incorporated into the thesaurus and, if so, where they should appear in the structure.

Terms occurring in user requests should be monitored in the same way. In particular, this is necessary to ensure that the vocabulary remains sufficiently specific to meet the information needs of the users.

UNIT 12. COMPUTER USE

Objectives

1. Show how computer processing can be used in the construction and maintenance of thesauri.

2. Demonstrate the important role that a machine-readable thesaurus can play in information retrieval activities.

Exhibit 65

Computers can play a very valuable role in the construction and exploitation of thesauri.

The possible application of computers to the gathering of terms was mentioned earlier (Exhibit 15). Terms can be derived from machine-readable data bases and ranked lists produced on the basis of frequency of occurrence. Even if terms are not derived from machine-readable sources, there is much to be said for putting them in machine-readable form to allow sorting and counting, and possibly to produce co-occurrence tables.

Because the construction of a conventional thesaurus is very largely an intellectual activity, computer processing of terms can only contribute to this task in relatively minor ways. For example, by sorting and printing all terms according to final words (welding, gas welding, shielded arc welding) or even by the roots of these words, the computer can perhaps aid in the identification of facets or hierarchies.

Once the thesaurus compiler has organized the vocabulary into facets and hierarchies, however, computer processing can be extremely valuable, if not absolutely essential. As an example, consider the partial hierarchy of Exhibit 8, where a simple notation is used to represent the hierarchical levels. Given such data, the computer can print a hierarchical display such as that shown in Exhibit 7, a permuted word display such as that of Exhibit 50, and (providing each hierarchy has been placed in a particular subject category), a categorized display such as that of Exhibit 48.

More importantly, however, a computer program can be used to generate a perfectly consistent alphabetical display (as in Exhibit 10) with all BT's, NT's, RT's and use references perfectly reciprocated. Furthermore, the thesaurus data in machine-readable form can be input to a photocomposition device so that the final printout of all displays can be of high typographic quality.

As well as producing a thesaurus, and presenting it in a number of output formats, computer programs can be used to maintain the thesaurus. Terms may be added at any time. The addition of a new term, in the appropriate way, will cause the term to be incorporated into the thesaurus structure in its correct position, and the necessary reciprocals to be generated. When a term is deleted from the thesaurus, the programs will check all the term reciprocals and delete all references to this term automatically. If the spelling of a term is changed, the programs will alter the spelling in all the appropriate places. If additional cross references are added to a term, the programs will automatically generate the necessary reciprocals for this reference.

In a similar way, the computer will make all needed thesaurus alterations when a scope note is changed, or when a term is moved from one subject category to another.

Exhibit 66

The computer can play a valuable role in generating a printed thesaurus. The machine-readable thesaurus can also have important functions to undertake in the operation of a retrieval system.

The machine-readable thesaurus is a fully authoritative record of the system vocabulary, and it is completely up-to-date. Let us call it the "master vocabulary file" or MVF. In addition to providing thesaurus printouts of various types, the MVF has the following functions in a computer-based retrieval system. It checks for consistency and acceptability of terms used by indexers and by searchers. Should an indexer or searcher use a term that is not recognized by the MVF, the input record or search input will be rejected with an appropriate notification to indexer or searcher.

In some systems the MVF may undertake certain automatic mapping activities. For example, the indexer or searcher may be allowed to use any of several synonyms recognized in the thesaurus, the MVF automatically substituting the preferred term. Only when a term used by indexer or searcher is completely unrecognized will the input be rejected.

The MVF maintains certain useful statistics that would be difficult or impossible to maintain manually. First and foremost, it maintains a record of the number of "postings" under each term; i.e., the number of times it has been used in indexing. Such data can be presented in printouts or displayed online. Information on the frequency with which a term has been used in indexing is valuable in the searching operation because it allows the searcher to estimate how many citations will be retrieved in response to a particular search strategy or, at the very least, show the maximum number of citations that could possibly be retrieved.

Statistics on frequency of assignment in indexing are important also in vocabulary control activites. A descriptor that has been used very infrequently in the past twelve months may be a good candidate for deletion from the vocabulary.

Statistics on frequency of use of terms in searching are also extremely valuable in vocabulary control activities (perhaps even more valuable than the statistics on indexing use); these statistics are less frequently used in automated systems although they are no more difficult to collect.

A record of the complete hierarchical structure of the vocabulary should appear on the MVF. This facilitates the conduct of generic searches. Only the parent term need be specified in searching, the descendants of it being substituted automatically from the MVF. Considering once more Exhibit 49, some means should exist for pulling into a search all terms from a particular hierarchy--e.g., all "vehicle" terms or all "motor vehicle" terms.

Using statistics on frequency of term assignment in indexing, the MVF may be used in the automatic optimization of a search strategy. Consider the Boolean strategy (A or B) and (L or M or N) and Y. To satisfy this search formula a citation must have been indexed under either term A or B and (L or M or N), and also under term Y. Suppose that Y is a very broad term and has been used 10,000 times, A and B have collectively been used 750 times, and L, M, N have collectively been used only 84

times. In searching the system it would be most efficient to read the shortest list of document numbers first and then to compare with the next shortest list, comparison with the largest list being left to the last step in the process. From postings statistics on the MVF, an efficient search program can automatically optimize the strategy; i.e., identify the least-heavily posted component, the next least-heavily posted, and so forth, ordering the strategy in ascending order of posting, and manipulating files in this sequence to economize on computer time.

The MVF may be used to generate automatically see and see also references in printed indexes. For example, when an entry is to appear in the printed index under EVAPORATION, the reference VAPORIZATION see EVAPORATION is generated automatically and will appear in the printed index. Likewise, if DEMODULATORS and MODULATORS are both used as headings in an issue of a printed index, because they are linked by RT indicators in the machine thesaurus, the printed see also references are generated automatically.

Exhibit 67

In an online information system, both indexer and searcher will need to consult an alphabetical display of the vocabulary. That is, on some occasions at least, they will need to verify that a particular term does exist in a data base. The capability of displaying a vocabulary alphabetically is fairly common in existing online systems. In response to a command such as NEIGHBOR or EXPAND, for any term entered, the system will display the term itself together with the terms that immediately precede and follow it in alphabetical sequence. To the right of each term appears a numeral indicating the number of postings for that term. Note that the term entered by the searcher has no postings data associated with it. This means that this particular term does not appear in the data base. Nevertheless, the system displays those terms that are closest to it alphabetically. This feature may help the searcher to recognize the term he really needs to use (in this case, perhaps, FLEXIBILITY) and it is a feature that will compensate for certain errors of misspelling.

It is clear that a vocabulary display capability of this type is an essential feature of an online system, whether the data base to be searched uses a controlled or an uncontrolled vocabulary. Not only can such a feature be used to verify the existence of a term, but it may lead the searcher to an alternative, more appropriate term than the one he began with, to the extent that alphabetical proximity will bring related terms together. In this example the searcher may decide that FLEXIBLE WALLS is the most appropriate term for his particular information need. The online system may give the searcher the capability of "paging" up or down the alphabetical display. He could, in fact, view the entire vocabulary of the system in this way, although this activity would be a most inefficient and costly use of online connect time.

It is important that, when a vocabulary is displayed online in this way, the user should be able to select from the display without the need to enter terms at the keyboard. Selection of terms from the list shown can be achieved by use of identifying line numbers. This facility saves the time of the searcher and reduces the probability of error.

Exhibit 68

A simple alphabetical listing of terms, while needed for various purposes, gives little assistance in the formulation of a comprehensive search strategy. The searcher also needs the capability of viewing terms that are more clearly semantically related. For any term entered at the terminal, the system should be able to display all the terms

that the vocabulary shows to be related paradigmatically and syntagmatically. There are a number of forms in which such a display could be presented. It could be a conventional thesaurus display showing, for the starting term, the term or terms that are broader hierarchically (BT), the term or terms that are narrower (NT), and the term or terms related in a nonhierarchical way (i.e., the related terms or RT's). Again, the searcher should be able to select any term from this display without having to enter it in full.

Better perhaps than the thesaurus display is one in the form of a hierarchical classification, such as that shown in Exhibit 7 or Exhibit 8.

Not only should the online system be able to display term hierarchies, but it should also allow the searcher to incorporate an entire hierarchy of terms into a search strategy by some simple command. In Exhibit 8, for example, one might do this by asking for all "L" terms. In the case of a display of the type shown in Exhibit 68, the complete hierarchy of terms can be incorporated by use of all appropriate numbers (in this example, perhaps, 1, 3-7).

UNIT 13. VOCABULARY FACTORS AFFECTING THE PERFORMANCE OF INFORMATION SYSTEMS

Objectives

1. Demonstrate the objectives of an efficient information retrieval system.
2. Define recall and precision.
3. Demonstrate the importance of vocabulary specificity.
4. Demonstrate the importance of adequate hierarchical structure.
5. Define false coordination and incorrect term relationship and show how these problems can be avoided.
6. Distinguish between pre-coordinate and post-coordinate systems.

Exhibit 69

In information retrieval operations one would like to pull from a data base those records that can be considered relevant to some information need. At the same time, one wants to avoid retrieval of any records that are not relevant.

The rectangle represents a hypothetical data base. Each "+" represents a record that is relevant to a particular information need and each "-" a record not relevant. In performing a search, then, one would like to retrieve as many as possible of the "+" items and as few as possible of the "-" ones.

The result of two searches are shown. In A, three relevant items were retrieved, and no irrelevant ones. This can be considered a search of high precision. At the same time, it is a search of low recall because most of the relevant items have been missed.

Search B, on the other hand, retrieves most of the relevant items. At the same time, however, it also brings out many that are not relevant.

Recall refers to the ability to retrieve relevant items and precision to the ability to retrieve only relevant items (or, put differently, the ability not to retrieve irrelevant ones).

In terms of recall and precision, the results of the two searches can be expressed statistically as proportions or ratios. In A, the recall ratio is only 3/10 (30%) but the precision ratio is perfect at 3/3 (100%). In B, the recall ratio is very high, 9/10 (90%), but precision has dropped to 9/15 (60%). The two sets of results are realistic in the sense that a very broad search will tend to retrieve many items (recall may be high but precision may be low) while a very narrow search will retrieve many fewer items (so precision may be high but recall low). To get high recall one may have to sacrifice precision and to get high precision one may have to sacrifice recall.

Exhibit 70

There are many factors that influence the performance of a retrieval system as measured in terms of recall and precision. The major categories are listed here. In a sense, the first is most important of all--if a system user makes a request that poorly reflects his real information need, it is unlikely that the results of a search will be very satisfactory, however good the indexing, search strategies and vocabulary may be.

Searching failures may occur because of indexing problems: an indexer may leave out a term that should have been used, may use one that is not really appropriate, or may use one that is not sufficiently specific.

The person searching the data base may search too generally or too specifically, may use an incorrect term, or may overlook important terms that should be incorporated.

A detailed discussion of these factors is outside the scope of this course. Factors affecting the performance of a retrieval system that are directly attributable to the vocabulary, however, require some explanation.

Exhibit 71

The specificity of the terms is the most important characteristic of a controlled vocabulary. The diagram illustrates three possible levels of specificity. If the specific term lemons exists in a thesaurus, articles on this fruit can be indexed precisely. If the specific term does not exist, these articles would have to be indexed much less precisely--under citrus fruit or even, perhaps, fruit.

Specificity controls the precision that can be achieved in searching. If one must retrieve all items in a data base indexed under citrus fruit when only items on lemons are wanted, the precision ratio of the search will presumably be quite low. The situation would be even worse if the term fruit was the most specific available in this hierarchy.

Specificity must be related to the needs of the users of the information service, which is why the principles of literary warrant and user warrant are so important. In a thesaurus on agriculture one will presumably need specific terms for individual fruits, whereas in another subject field--perhaps economics or trade--the broad term fruit may be quite satisfactory. If in doubt, it is better to be too specific than not specific enough.

It should also be recognized that a nonspecific vocabulary, while it makes it difficult to achieve high precision, will tend always to give a high recall. If we are forced to retrieve everything on fruit when only items on lemons are wanted, recall may well be higher than would be the case if the specific term existed. For example, some "lemons" items may not have been indexed under the specific term because of indexing error and some of the more general reports indexed under citrus fruit may contain substantial information on lemons.

Exhibit 72

A second important requirement of a controlled vocabulary is that its structure must be such as to provide the maximum possible aid to indexers and searchers.

The diagram shows a very incomplete set of terms relating to pollution. Terms are linked by arrows: solid arrows reflect hierarchical relationships and broken arrows reflect relationships that cut across hierarchies.

The thesaurus must explicitly reveal all of the relationships depicted (and any others that seem useful) by its BT/NT/RT structure or other means. By so doing, it helps the indexer choose the most appropriate term to represent a particular concept. For example, it leads him from water pollution to thermal pollution, allowing him to index more correctly a report on "destruction of life in streams through heated discharges".

It serves the same function for the searcher and, in addition, serves one of even greater importance: it shows all of the terms needed for a comprehensive search on some topic. To undertake a broad search on pollution, the thesaurus tells us, we will need to include terms for various kinds of pollution and perhaps also bring in some terms from other hierarchies: pollutants, wastes, and so on.

If the structure of the thesaurus is imperfect or inadequate, this may lead to errors in indexing or in the performance of a search.

Exhibit 73

Precision failures may be caused by the fact that the vocabulary, or the way it is used, lacks explicit syntax. To understand this, one must distinguish between pre-coordinate systems and post-coordinate systems.

Consider a hypothetical report indexed under the terms listed. It deals with some aspects of current trade in Brazil and, in particular, its import/export relations with Argentina and Mexico.

Clearly, some of the relationships implied by the list are "valid" while others are "spurious". For example, the report discusses Brazilian export licenses but not Argentinian export licenses, and there is no direct relationship at all between Mexico and Argentina.

Exhibit 74

Here, several of the terms have been combined to form entries as these might appear in a conventional card catalogue or printed index. Spurious relationships are avoided--only terms directly related are brought together in an entry and the sequence of the terms helps to make the relationship explicit and unambiguous (e.g., Brazil is exporting to Argentina, not Argentina to Brazil). Nevertheless, different interpretations are still possible. To avoid these completely, more explicit syntax may have to be introduced.

This exemplifies a pre-coordinate index: terms are coordinated (combined) explicitly and these combinations are built into the index. This has the advantage of reducing ambiguities but brings its own disadvantages.

Exhibit 75

In a post-coordinate system, on the other hand, all of the terms assigned to a document by an indexer are tied in some way to the bibliographic record for that document but the terms are not combined into groups as they were in Exhibit 74. In fact, in this type of system the document hypothesized could be retrieved on any

combination--of two terms, three terms, four terms, and so on. Some of the possible relationships among the eight terms assigned are illustrated in this diagram by connecting lines.

Exhibit 76

A false coordination occurs when the terms that cause a document to be retrieved are not directly related in that document. For example, based on the indexing shown in Exhibit 73, this record would be retrieved in response to a search on exports from Argentina to Mexico. In point of fact, no relationship between Argentina and Mexico is discussed in the report; i.e., the two terms are quite unrelated.

False coordinations of this kind can be avoided. The most obvious way is to put the index terms into groups, as shown, so that all terms in a group are directly related. Some number or letter is used to identify each group and these symbols are used to maintain the separate identity of the groups within a data base. The device is usually referred to as a link because related terms are linked together.

Through the linking shown, the false coordination mentioned earlier can be avoided, since Mexico and Argentina appear in different groups. Many other possible false coordinations are also avoided--any of the trade terms in group 3 combined with Mexico or with Argentina.

Exhibit 77

A more complex situation, not readily avoided by links, is the incorrect term relationship. Even with the linking illustrated in Exhibit 76, this document could be retrieved in a search on exports from Argentina to Brazil. Instead it deals with the reverse situation: exports from Brazil to Argentina (Argentinian imports). Again, the situation can be solved--this time by the use of role indicators. In the example, two simple "roles" have been, as it were, added to the vocabulary. By using these as modifiers of the descriptors in the bibliographic record, we are able to indicate that Brazil is the exporter and Argentina the importer.

A relatively small number of role indicators, perhaps 10 or 12, expressing rather basic relationships, will solve most of these types of problems when they are selected to meet the needs of a particular subject field.

Exhibit 78

In a vocabulary consisting mostly of single words, false coordinations and incorrect term relationships may occur fairly frequently. They are less likely in a vocabulary that uses a sensible level of compounding (see Unit 7). For example, if a document is indexed under the term women and the term poetry it could deal with several possible topics: women represented in poetry, women as readers of poetry, women as writers of poetry, and so on. Similarly, computers and design could refer to design of computers or to design of something else (e.g., airplanes) using computers as a tool.

Many of these ambiguities are resolved by more compounding in the vocabulary; e.g., by inclusion of terms such as women in poetry, computer design, airplane design and computer-assisted design.

Another possible solution involves the use of subheadings. A relatively small number of carefully chosen subheadings, incorporated as a special component of the vocabulary, may avoid many possible ambiguities--as in the computers/design, airplanes/design example. In a "trade" information centre, exports and imports would be good candidates for use as subheadings. Thus, Brazil/exports and Argentina/imports provide an unambiguous representation.

It must be pointed out, however, that not all of these solutions are compatible with the international guidelines.

Exhibit 79

The use of role indicators or of subheadings in effect increases the size and specificity of the vocabulary. Instead of having a single term "Brazil", we now have several: Brazil in role A, in role B, or Brazil/imports, Brazil/exports, and so on.

Subheadings have obvious advantages over role indicators: they are "natural language" rather than symbolic and can be used to subdivide entries in printed indexes.

It must be recognized that the use of links and roles, in particular, adds complexity to indexing and searching and therefore increases costs. This applies to a lesser extent to a simple set of subheadings. Moreover, the example used as our original illustration (Exhibit 73) is an extreme one. Most indexing will be much less ambiguous than this. While false coordinations and incorrect term relationships will occur in all post-coordinate systems, they should be sufficiently infrequent to be "tolerable" without the use of additional devices, especially if the vocabulary has a reasonable level of compounding.

UNIT 14. NATURAL LANGUAGE SYSTEMS

Objectives

1. Illustrate again the major problems involved in operating a system without vocabulary control.

2. Illustrate techniques that can be used to compensate for lack of vocabulary control in computer-based systems.

3. Introduce and demonstrate the postcontrolled vocabulary.

Exhibit 80

This course deals with controlled vocabularies, most particularly the thesaurus. In a purely manual environment, it is virtually impossible to operate a retrieval system successfully without vocabulary control. When computers are used, however, it becomes more feasible to function without vocabulary control (i.e., operate with "natural language"). For example, the words appearing in titles and abstracts, stored in a data base, may be used in place of index terms to represent the subject matter of documents. These words can be searched in any combinations. Thus, one can ask the computer to locate any abstracts containing the word "corrosion" and the word "copper".

The problem with a completely natural-language system is that it gives no help to the searcher. To retrieve all information on "wastes", for example, we may have to think not only of this word but of such others as "sewage", "discharges" and "effluents". A good thesaurus would link these terms in some way or even use one term to stand for the whole group.

Likewise, a thesaurus would show us the many different terms that may be needed for a comprehensive search on "magnetism".

The user of a natural-language system must compensate for lack of vocabulary control by using ingenuity.

Exhibit 81

Nevertheless, there are some techniques available for searching natural-language systems that help to compensate for lack of a controlled vocabulary. The most powerful is truncation or word fragment searching--the ability to search on parts of words as well as on complete words.

Three major kinds of truncation exist:

1. Searching on the initial characters in a word (right truncation): all words beginning with "magnet" could be retrieved.

2. Searching on the final characters (left truncation): all words ending with "magnet" could be retrieved.

3. Searching on the character string "magnet" wherever it occurs in a word (left and right truncation combined). In addition to retrieving words beginning and ending with "magnet", this would also retrieve words with "magnet" in the middle.

In this example, then, word fragment searching would be very powerful because the "magnet" fragment would be likely to retrieve most documents having anything to do with magnetism in any form.

Exhibit 82

It is obvious that false coordinations can easily occur in the searching of abstracts; the longer the abstracts, the greater the probability. Some systems offer the user a chance to reduce this probability by allowing him to specify how close the two words should be in the text (e.g., no more than 10 words intervening). The closer two words appear in an abstract, the greater the chance that they will be directly related. All systems should at least allow "adjacency"--e.g., the word "copper" followed immediately by "alloys".

Exhibit 83

It is possible to have a retrieval system that is a "hybrid" in the sense that it uses some controlled terms as well as natural language. For example, some data bases incorporate humanly assigned descriptors and humanly assigned (uncontrolled) keywords while others combine descriptors with the text of abstracts.

It is also possible to superimpose on a natural-language system a kind of controlled vocabulary "superstructure"--a rather small set of broad codes assigned by indexers. This helps to reduce ambiguity (e.g., when the word "strikes" occurs in conjunction with the "labor" code it is likely to mean something different than when it occurs with the "warfare" code) and greatly facilitates the conduct of broad searches (e.g., all aspects of warfare related to a particular region).

Exhibit 84

To aid the searching of natural-language data bases, it is also possible to build (and store in machine-readable form) tables of synonyms, near-synonyms or other conceptually related terms. These can include fragments as well as complete words. Two examples are given here, where the * indicates truncation. The first table is designed to retrieve items having something to do with "waste" while the second is likely to retrieve most abstracts that deal in any way with antibiotics.

Such tables are given names and perhaps identifying numbers. They could be displayed online or incorporated into a search strategy in much the same way that a complete hierarchy might be incorporated from a thesaurus.

It is obvious that such tables serve some of the same purposes that thesauri are designed to serve. In fact, they can be looked upon as a rather different type of "controlled vocabulary"--a postcontrolled vocabulary, one intended to assist the searching of data bases but not to impose any control over the terminology input into the system.

As the cost of human intellectual processing continues to escalate relative to the cost of computer processing, it is likely that information services may move increasingly toward natural language and away from conventional controlled vocabularies. As this occurs, more attention will be paid to the construction of postcontrolled vocabularies. A rather different type of thesaurus, much less structured, may emerge.

EXHIBITS

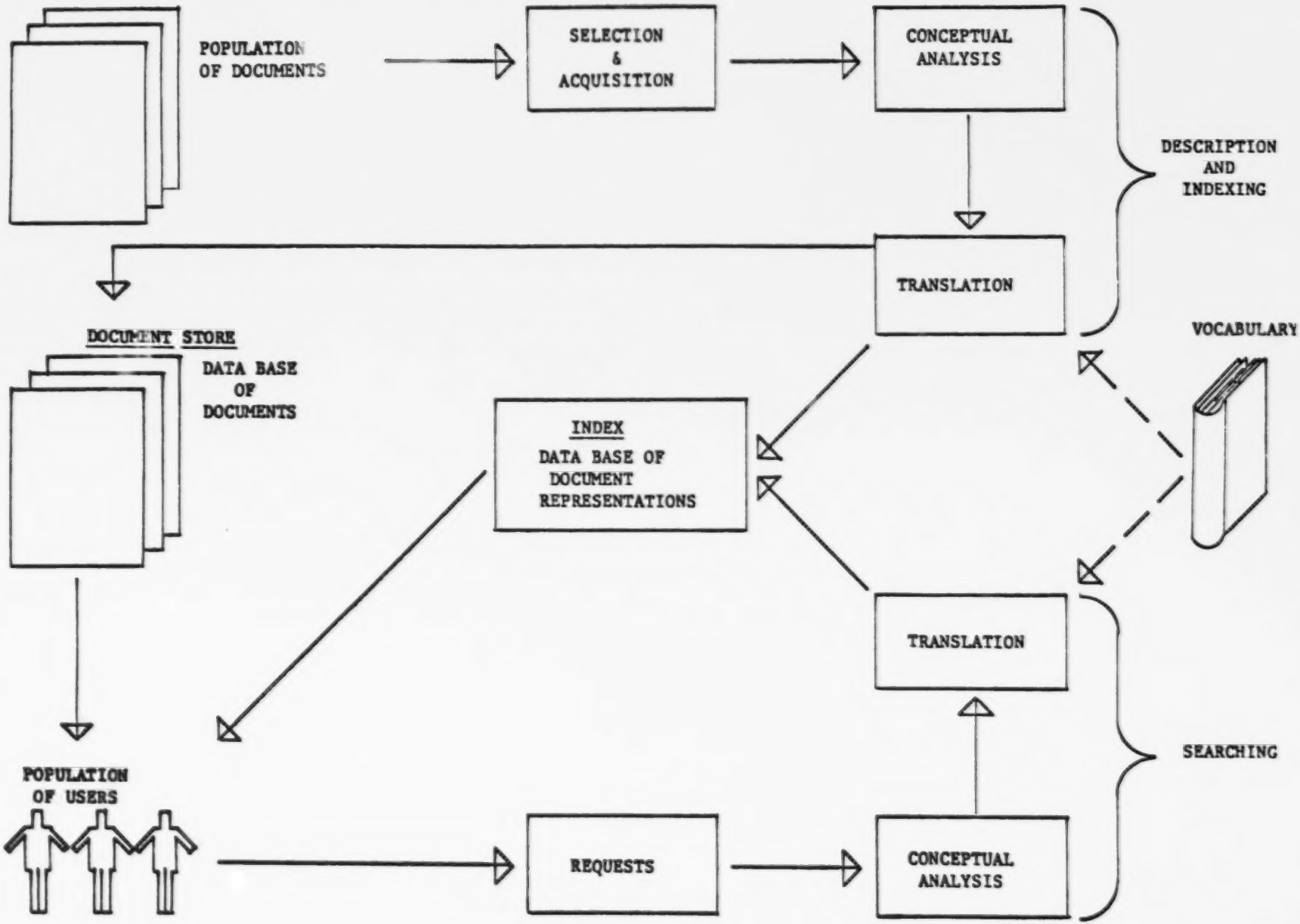

POPULATION OF DOCUMENTS
SELECTION & ACQUISITION
CONCEPTUAL ANALYSIS
DESCRIPTION AND INDEXING
TRANSLATION
DOCUMENT STORE
DATA BASE OF DOCUMENTS
VOCABULARY
INDEX
DATA BASE OF DOCUMENT REPRESENTATIONS
TRANSLATION
SEARCHING
POPULATION OF USERS
REQUESTS
CONCEPTUAL ANALYSIS

Exhibit 1

Brittle fracture	Fatigue
Brittleness	Flaws
Corrosion	Fracture
Cracking	Hydrogen embrittlement
Creep	Rupture
Creep rupture	Strength
Defects	Stress
Deterioration	Stress corrosion
Ductile fracture	Stress rupture
Ductility	Tensile strength
Failure	

Exhibit 2

(a) Defects ——→ Flaws

(b) Cracking (fracture; e.g., of metals)

Cracking (splitting of molecules as used in chemical processing; e.g., in petroleum technology)

Fatigue (of humans)

Fatigue (of materials)

(c)

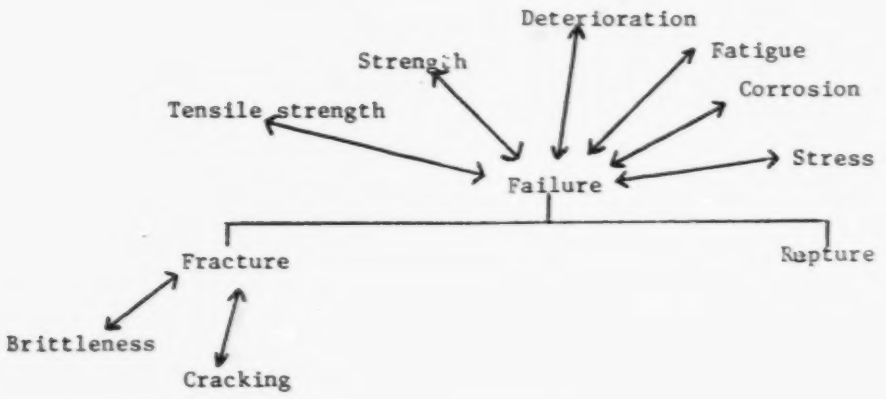


Exhibit 3

1. Promote consistent representation
   - at input (indexing)
   - at output (searching)

   by:
   - controlling synonyms
   - distinguishing among homographs

2. Facilitate comprehensive searches by linking semantically related terms:
   - genus/species relationship
   - other types

Exhibit 4

Complementary components

(a) alphabetical

(b) classified

May be
- overtly classified with alphabetical index
- overtly alphabetical with covert classification built in

Exhibit 5

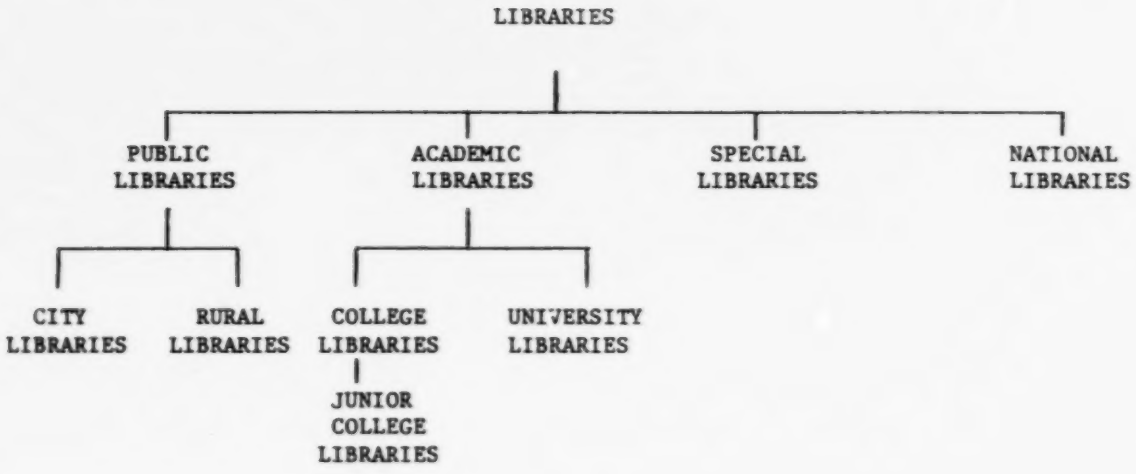


Exhibit 6

Libraries

- Public libraries
  - City libraries
  - Rural libraries
- Academic libraries
  - College libraries
    - Junior college libraries
  - University libraries
- Special libraries
- National libraries

Exhibit 7

L	Libraries
Lc	Public libraries
Lcc	City libraries
Lcr	Rural libraries
Le	Academic libraries
Lec	College libraries
Lecj	Junior college libraries
Leu	University libraries
Lg	Special libraries
Li	National libraries

Exhibit 8

Academic libraries	Le
City libraries	Lcc
College libraries	Lec
Junior college libraries	Lecj
Libraries	L
National libraries	Li
Public libraries	Lc
Rural libraries	Lcr
Special libraries	Lg
University libraries	Leu

Exhibit 9

Academic libraries

BT Libraries

NT College libraries
University libraries

City libraries

BT Public libraries

College libraries

BT Academic libraries

NT Junior college libraries

Junior college libraries

BT College libraries

Libraries

NT Academic libraries
National libraries
Public libraries
Special libraries

National libraries

BT Libraries

Public libraries

BT Libraries

NT City libraries
Rural libraries

Special libraries

BT Libraries

University libraries

BT Academic libraries

Exhibit 10

(a) Libraries

Public libraries

City libraries → City government
Rural libraries

Academic libraries → Higher education

College libraries
University libraries

(b) Academic libraries

BT Libraries

NT College libraries
University libraries

RT Higher education

City libraries

BT Public libraries

RT City government

Exhibit 11

1. Adopt existing thesaurus
2. Modify (e.g., expand) existing thesaurus
3. Translate existing thesaurus
4. Build new thesaurus

Exhibit 12

1. Gathering terms

2. Organizing terms

3. Producing final structure

4. Printout and display

Exhibit 13

1. "Theoretical", top-down approach (deductive)

2. Empirical, bottom-up approach (inductive)

   Literary warrant
   User warrant
   Combined literary and user warrant

Exhibit 14

Literary warrant

Reference works: dictionaries, glossaries, encyclopedias, handbooks, textbooks, indexes

Abstracts

Printouts from data bases

By term occurrence

——— 1172
——— 1107
——— 1049
——— 1002

By term co-occurrence

A ——— Q
——— R
——— S

Exhibit 15

User warrant

1. Search requests

2. Questionnaire to sample of users

   Job description
   Research interests
   Titles/abstracts of publications

3. Users select terms from the literature

Exhibit 16

Experience in thesaurus construction
↓
Guidelines
↓
National standards
↓
International standards

Exhibit 17

1. Collect terms

2. Identify major categories (classes) of terms

3. Establish hierarchies within these categories

4. Establish useful relationships among terms from different hierarchies

5. Convert to alphabetical organization

Exhibit 18

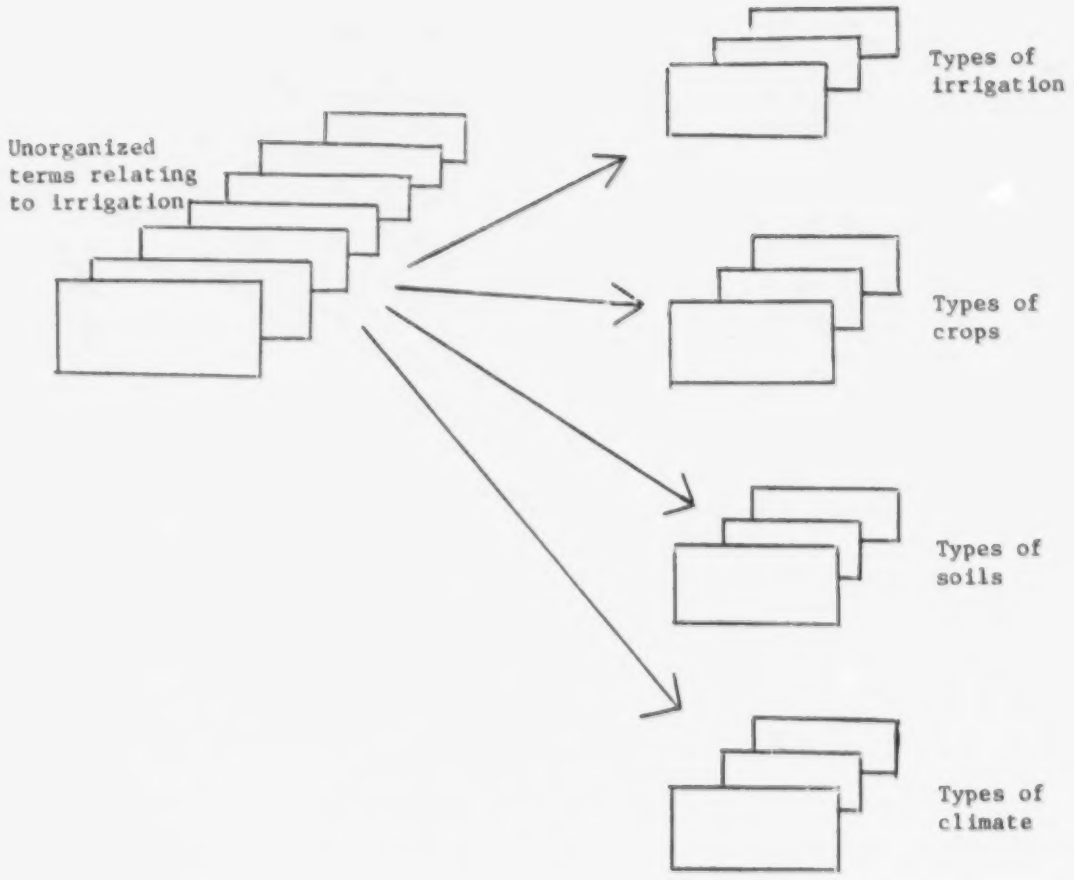

Unorganized terms relating to irrigation
Types of irrigation
Types of crops
Types of soils
Types of climate

Exhibit 19

Crops

- Fruits
  - Citrus fruits
    - Oranges
    - Lemons
    - Limes
    - Grapefruit
  - Vegetables
  - Grains

Exhibit 20

Citrus fruits
BT Fruits
NT Grapefruit
Lemons
Limes
Oranges

Crops
NT Fruits

Fruits
BT Crops
NT Citrus fruits

Grains
BT Crops

Grapefruit
BT Citrus fruits

Lemons
BT Citrus fruits

Limes
BT Citrus fruits

Oranges
BT Citrus fruits

Vegetables
BT Crops

Exhibit 21

Crops

NT Citrus fruits
Fruits
Grains
Grapefruit
Lemons
Limes
Oranges
Vegetables

Crops

NT Fruits
Citrus fruits
Grapefruit
Lemons
Limes
Oranges
Grains
Vegetables

Exhibit 22

Crops

NT1 Fruits

NT2 Citrus fruits

NT3 Grapefruit
Lemons
Limes
Oranges

NT1 Grains

NT1 Vegetables

Exhibit 23

1. Generic Relationship

2. Partitive Relationship

3. Instance Relationship

Exhibit 24

Cardiovascular System

Heart

Aorta

South America

Venezuela

Caracas

Exhibit 25

Ships

NTG Freighters
Passenger Ships
Warships

NTP Bows
Decks
Hulls

Exhibit 26

Cathedrals

NT Coventry Cathedral
Durham Cathedral
Westminster Cathedral

Exhibit 27

BEST COPY AVAILABLE

ARCHITECTURE

[by style]

Baroque

Gothic

Romanesque

[by country]

American

French

German

[by application]

Domestic

Ecclesiastical

Naval

Exhibit 28

ARCHITECTURE

NT American Architecture

Baroque Architecture

Domestic Architecture

Ecclesiastical Architecture

French Architecture

Gothic Architecture

etc. etc.

Exhibit 29

ARCHITECTURE

NT (by application)

Domestic Architecture

Ecclesiastical Architecture

Naval Architecture

NT (by style)

Baroque Architecture

Gothic Architecture

Romanesque Architecture

Exhibit 30

DIAMONDS

BT CUTTING TOOLS

DIAMONDS

BT PRECIOUS STONES

Exhibit 31

ORANGES

TT Crops

BT Citrus fruit

Exhibit 32

ARCHITECTURE

NT Domestic Architecture

Ecclesiastical Architecture → Churches, Monasteries, Iconography

Naval Architecture

ECCLESIASTICAL ARCHITECTURE

BT Architecture

RT Churches

Iconography

Monasteries

Exhibit 33

Thing/application

Water supply

RT Irrigation

Result/cause

Aftershocks

RT Earthquakes

Thing/property

Gifted children

RT Intelligence

Raw material/product

Bauxite

RT Aluminium

Complementary activities

Buying

RT Selling

Opposites

Life

RT Death

Activity/property

Cutting

RT Machinability

Activity/agent

Smoking

RT Tobacco

Activity/product

Weaving

RT Cloth

Whole/part

Airplanes

RT Wings

Exhibit 34

Nouns or noun phrases

Gravity (noun)

Specific Gravity (adjectival phrase)

Centre of Gravity (prepositional phrase)

Direct, not inverted

Comic Books	Not Books, Comic
Picture Books	Books, Picture

Exhibit 35

Plural	Singular
Nouns of quantity ("count nouns")	Nouns of volume
How many?	How much?
Dogs	Corn
Mice	Flour
Railroads	Paper

Exhibit 36

(Library) Cooperation

(Library) Architecture

(Library) Buildings

Exhibit 37

1. Automobile Engine Noise

2. Automobile Engines Noise

3. Automobiles Engine Noise

4. Automobiles Engines Noise

Exhibit 38

Split Compound	Retain Compound
1. Traffic Noise into Traffic Noise	4. Acidity of Paper
2. Children's Picture Books into Children's Books Picture Books	5. Tiger Lilies 6. Plate Glass 7. Clothes Horses 8. Monroe Doctrine 9. Pressure Vessels
3. Turbine Blade Corrosion into Turbine Blades Corrosion	

Exhibit 39

Entry Vocabulary

1. Synonyms and near-synonyms
2. Terms considered unnecessarily specific in context of thesaurus

Exhibit 40

1. Cars use Automobiles
2. GNP use Gross National Product
3. Carousels use Merry-go-rounds
4. Acoustics use Sound
5. Motion, equations of use Equations of motion
6. Inelasticity use Elasticity
7. Domestic architecture use Architecture and Houses
8. White noise use Noise

Exhibit 41

Cars use Automobiles

Automobiles UF Cars

Domestic architecture use Architecture and Houses

Architecture

UF† Domestic architecture

Houses

UF† Domestic architecture

Exhibit 42

PERMAFROST

SN Earth or rock that is perennially or permanently frozen.

CATS

SN Use only for discussions on the cat family in general, including wild and domesticated animals. Whenever appropriate, prefer the more specific terms, e.g., DOMESTIC CATS, LIONS, TIGERS.

AGING

SN The obsolescence of library materials, usually measured in terms of declining use with age. For physical decline of library materials use DETERIORATION.

Exhibit 43

Tanks (containers)	Beams (electronics)
Tanks (weapons)	Beams (structural elements)

Exhibit 44

Identifiers

1. Names of weapons and other equipment
2. Personal names (e.g., authors, historical characters)
3. Geographic names
4. Chemical compounds
5. Institutions

Exhibit 45

POSITION FINDING INSTRUMENTS

SN Instruments used to locate aircraft, ships or other objects in relation to a specific reference point or points

UF Position indicators

BT INSTRUMENTS

NT COMPASSES

GROUND POSITION INDICATORS

PLAN POSITION INDICATORS

SEXTANTS

TARGET POSITION INDICATORS

RT DETECTION

TRACKING

Exhibit 46

IRON AND STEEL INDUSTRY
05.03.00 DBLMI
UF STEEL WORKS
BT MINERAL INDUSTRY
RT BLAST FURNACES
CAST IRON
COKE INDUSTRY
DESULPHURATION
IRON
IRON AND STEEL WORKERS
STEEL
TT INDUSTRY

IRON AND STEEL WORKERS
12.05.00 D
BT INDUSTRIAL WORKERS
RT IRON AND STEEL INDUSTRY
TT OCCUPATIONS

IRON ORE
07.02.02 03.09.00 DMLIU
BT MINERALS
RT IRON
TT MINERALS

IRRADIATION
16.08.00 DMS
BT RADIATION
RADIOBIOLOGY
RT GAMMA RADIATION
IONIZING RADIATION
RADIATION EFFECTS
RADIOACTIVITY
TT SCIENCE AND TECHNOLOGY

IRRIGATION
04.02.01 DBMULIS
UF IRRIGATION ENGINEERING
BT HYDRAULIC ENGINEERING
NT TRICKLE IRRIGATION
RT DAMS
DRAINAGE
EVAPORATION (METEOROLOGY)
IRRIGATION CANALS
IRRIGATION FARMING
IRRIGATION RESEARCH
SPRAYING EQUIPMENT
WATER STORAGE
TT HYDRAULICS
SCIENCE AND TECHNOLOGY

IRRIGATION CANALS
04.02.01 SU
RT CANALS
IRRIGATION

IRRIGATION ENGINEERING
04.02.01 D
USE IRRIGATION

IRRIGATION FARMING
04.02.01 DF
BT CULTIVATION SYSTEMS
RT IRRIGATION
TT CULTIVATION SYSTEMS

IRRIGATION LAW
04.01.01 D
BT WATER LAW
RT WATER RIGHTS
TT LAW

IRRIGATION RESEARCH
04.01.02 D
BT AGRICULTURAL RESEARCH
RT IRRIGATION
TT SCIENTIFIC RESEARCH

ISLAM
15.03.00 DMULS
UF MOHAMMEDANISM
BT RELIGION
NT ISLAM AND ECONOMICS
ISLAM AND PHILOSOPHY
ISLAM AND POLITICS
ISLAM AND SOCIAL PROBLEMS
ISLAM AND SOCIALISM
ISLAM AND STATE

ISLAM (cont.)
RT ISLAMIC ART
ISLAMIC CIVILIZATION
ISLAMIC COUNTRIES
ISLAMIC EPISTEMOLOGY
ISLAMIC ETHICS
ISLAMIC PRAYERS
ISLAMIC SOCIOLOGY
ISLAMIC THEOLOGY
KORAN
PANISLAMISM
WOMEN IN THE KORAN
TT RELIGION

ISLAM AND ART
15.03.00 D
USE ISLAMIC ART

ISLAM AND ECONOMICS
15.03.00 D
BT ISLAM
TT RELIGION

ISLAM AND PHILOSOPHY
15.03.00 D
BT ISLAM
TT RELIGION

ISLAM AND POLITICS
15.03.00 D
BT ISLAM
RT ISLAM AND SOCIALISM
TT RELIGION

ISLAM AND SOCIAL PROBLEMS
15.03.00 D
BT ISLAM
RT SOCIAL PROBLEMS
TT RELIGION

ISLAM AND SOCIALISM
15.03.00 D
UF SOCIALISM AND ISLAM
BT ISLAM
RT ISLAM AND POLITICS
TT RELIGION

ISLAM AND STATE
15.03.00 D
BT ISLAM
RT CHURCH AND STATE
TT RELIGION

ISLAMIC ANTIQUITIES
15.01.02 D
RT ARCHAEOLOGY
CHRISTIAN ANTIQUITIES

ISLAMIC ART
15.03.00 U
UF ISLAM AND ART
RT ISLAM

ISLAMIC CIVILIZATION
15.01.02 D
BT CIVILIZATION
RT ARABIC LITERATURE
ISLAM
TT CIVILIZATION

ISLAMIC CONFERENCE
17.07.01 D
USE ORGANIZATION OF THE ISLAMIC CONFERENCE

ISLAMIC COUNTRIES
15.03.00 D
RT ARAB COUNTRIES
ISLAM

ISLAMIC COURTS
01.07.00 D
BT COURTS
RT ISLAMIC LAW
TT JUDICIAL SYSTEM

ISLAMIC EPISTEMOLOGY
15.03.00 D
UF KNOWLEDGE, THEORY OF (ISLAM)
BT PHILOSOPHY
RT ISLAM
THEORY OF KNOWLEDGE
TT HUMANITIES

ISLAMIC ETHICS
15.03.00 D
BT ETHICS
RT ISLAM
TT ETHICS

ISLAMIC LAW
01.07.00 D
UF MOHAMMEDAN LAW
BT LAW
NT FATWAS
RT ISLAMIC COURTS
KORAN
TT LAW

ISLAMIC PRAYERS
15.03.00 D
RT ISLAM
KORAN

ISLAMIC SOCIOLOGY
14.05.01 D
BT SOCIOLOGY
RT ISLAM
TT SOCIAL SCIENCES

ISLAMIC THEOLOGY
15.03.00 D
BT THEOLOGY
RT ISLAM
TT THEOLOGY

ISLANDS
03.01.03 01.07.03 DMS
SN Includes concept "regime of islands" in the law of the sea
BT LANDFORMS
NT ARCHIPELAGOES
BARRIER ISLANDS
CORAL REEFS AND ISLANDS
DEVELOPING ISLAND COUNTRIES
TT LANDFORMS

ISOTOPE SEPARATION
16.08.00 DS
BT NUCLEAR TECHNOLOGY
RADIOCHEMISTRY
NT RADIOACTIVE TRACERS
RT ISOTOPES
RADIOACTIVE MATERIALS
RADIOISOTOPES
TT SCIENCE AND TECHNOLOGY

ISOTOPES
16.08.00 DBMUFS
NT RADIOISOTOPES
RT ATOMS
HYDROGEN
ISOTOPE SEPARATION
METALS

ISRAEL
17.05.00 D
BT MEDITERRANEAN COUNTRIES
RT ISRAEL-ARAB WAR (1948-1949)
ISRAEL-ARAB WAR (1967)
ISRAEL-ARAB WAR (1973)
ISRAELI
JEWISH LAW
JORDAN RIVER QUESTION
MIDDLE EAST
MIDDLE EAST SITUATION
PALESTINE QUESTION
PALESTINIANS
ZIONISM
TT MEDITERRANEAN COUNTRIES

ISRAEL AND THE DIASPORA
01.01.00 D
USE ZIONISM

ISRAEL-ARAB WAR (1948-1949)
01.06.02 D
SN Used for works dealing with the war itself and its immediate results. Otherwise use "Palestine question"
RT ARAB COUNTRIES
ISRAEL
PALESTINE QUESTION

Exhibit 47

70

BEST COPY AVAILABLE

04.02.01 (cont.)

COMPOSTING
  SN Controlled process of degrading organic matter by microorganisms
CROP DIVERSIFICATION
CROP MANAGEMENT
CROP ROTATION
CROP SUBSTITUTION
  SN Replacement of illicit narcotic raw materials with non-narcotic crops
CROP YIELDS
CROP ZONES
CROPPING SYSTEMS
  USE CULTIVATION SYSTEMS
CROPS
CULTIVATION SYSTEMS
DRY FARMING
  SN A system of producing crops in arid and semi-arid regions without the use of irrigation
FERTILIZER USE
FERTILIZERS
FIELD CROPS
FLORICULTURE
  SN Ornamental gardening
FORAGE CROPS
  SN Crops cultivated to provide leaves and stalks for feeding livestock. For forage plants in general, use "Pasture and forage"
FORAGE PLANTS
  USE FORAGE CROPS
  PASTURE AND FORAGE
FRUIT CULTURE
GREENHOUSES
HORTICULTURE
  SN Cultivation of gardens or orchards, including the growing of vegetables, fruits, flowers and ornamental shrubs and trees
HOT HOUSES
  USE GREENHOUSES
IRRIGATION
IRRIGATION CANALS
IRRIGATION ENGINEERING
  USE IRRIGATION
IRRIGATION FARMING
NITROGEN FERTILIZERS
  USE FERTILIZERS
NUCLEAR AGRICULTURE
ORGANIC FARMING
ORGANIC GARDENING
  USE ORGANIC FARMING
PLANT BREEDING
PLANT DISEASES
PLANT GENETICS
PLANT NAMES
PLANT NUTRITION
PLANT PATHOLOGY
  USE PLANT DISEASES
PLANT PROTECTION
RADIOISOTOPES IN AGRICULTURE
  USE NUCLEAR AGRICULTURE
ROTATION OF CROPS
  USE CROP ROTATION
SERICULTURE
SHIFTING CULTIVATION
SUGAR GROWING
  SN Refers to problems related to growing sugar crops
TRICKLE IRRIGATION
TROPICAL CROPS
  SN Use with respect to the cultivation of tropical crops. For tropical vegetation in general, use "tropical plants"
VITICULTURE

04.02.02
CROPS

ABACA
APPLES
APRICOTS
ARTICHOKES
ASPARAGUS
BAGASSE
  SN Fibrous residue resulting from the extraction of juice from the sugar cane
BAMBOO
BANANAS
BARLEY
BEANS
BRUSSELS SPROUTS
CACAO
  USE COCOA
CARDAMON
CARROTS
CASSAVA
CASTOR BEANS
CELERY
CEREALS
  USE GRAINS
CHERRIES
CINNAMON
CITRUS FRUITS
CLOVES
COARSE GRAINS
COCOA
COCONUT PRODUCTS
COFFEE
COIR
COPRA
COTTON
DATE PALMS
FLAX
FLOWERS
FRUIT
FRUIT TREES
GARLIC
GRAINS
GRAPES
GRASSES
GROUNDNUTS
HARD FIBRES
HEMP
  SN Use with respect to cannabis sativa as a fibre plant as well as the fibre obtained therefrom. For cannabis sativa as a drug plant, use "cannabis"

HENEQUEN
HORSE RADISH
JUTE
KENAF
LEEKS
LEGUMES
MAIZE
MILLET
MOLASSES
NUTMEG
NUTS
OATS
OILSEED PLANTS
OILSEEDS
OLIVES
ONIONS
ORANGES
PEACHES
PEANUTS
  USE GROUNDNUTS
PEARS
PEAS
PEPPER
PLANTS
PLUMS
POISONOUS PLANTS
POTATOES
PULSES
  USE LEGUMES
RAISINS
RAMIE
RICE
ROOT CROPS
RYE
SEEDS
SISAL
SORGHUM
SOYBEANS
SPICES
SUGAR
  SN Use only for discussion of sugar as a commodity, or, in agricultural context when sugar beets or sugar cane cannot be identified
SUGAR BEETS
SUGAR CANE
SWEET POTATOES
TABLE OLIVES
  USE OLIVES
TEA
TOBACCO
TRITICALE
TROPICAL PLANTS
  SN Use with respect to tropical vegetation in general. For cultivation of crops use "tropical crops"

TUBERS
  USE ROOT CROPS
TURMERIC
VEGETABLES
WHEAT
WINGED BEANS
YAMS

04.03.00
LIVESTOCK

AFRICAN ANIMAL TRYPANOSOMIASIS
AFTOSA FEVER
ANIMAL DISEASES
ANIMAL FEED
  SN Use with respect to processed or prepared food for livestock. For grasses or the leaves and stalks of plants on which livestock graze, use "pasture and forage"
ANIMAL FEED INDUSTRY
ANIMAL GENETICS
ANIMAL HEALTH
ANIMAL HUSBANDRY
ANIMAL INDUSTRY
  USE LIVESTOCK INDUSTRY
ANIMAL NUTRITION
ANIMAL PRODUCTION
  USE LIVESTOCK INDUSTRY
ANIMAL PRODUCTS
APICULTURE
BEE CULTURE
  USE APICULTURE
BEEF
BREEDING STOCK
BUTTER
CAMELS
CATTLE
CATTLE FEED
CHEESE
COMMUNICABLE DISEASES IN ANIMALS
  USE ANIMAL DISEASES
DAIRY FARMS
DAIRY PRODUCTS
DAIRYING
DRAFT ANIMALS
  USE DRAUGHT ANIMALS
DRAUGHT ANIMALS
EGGS
FEED SUPPLY
FOOT AND MOUTH DISEASE
  USE AFTOSA FEVER
FORAGE PLANTS
  USE FORAGE CROPS
  PASTURE AND FORAGE
GRAZING
HONEY
HORSES
LAMB
  SN Meat product
LIVESTOCK
LIVESTOCK INDUSTRY
MEAT
MERINO SHEEP
  USE SHEEP
MILK
PASTURE AND FORAGE
  SN Leaves and stalks of plants as well as grasses on which animals graze. For processed food for livestock, use "animal feed". For plants grown for the purpose of providing leaves and stalks for feeding livestock, use "forage crops"
POULTRY
RANGE MANAGEMENT
REINDEER BREEDING
SHEEP
STOCK AND STOCK-BREEDING
  USE LIVESTOCK INDUSTRY
SWINE
VEAL
VETERINARY HYGIENE
  USE VETERINARY MEDICINE
VETERINARY MEDICINE

04.04.00
FISHERIES

AQUACULTURE
COD FISHERIES
DEEP SEA FISHING
FISH
FISH CULTURE
  SN Refers to inland water fish culture

BEST COPY AVAILABLE

UNEMPLOYMENT
. SEASONAL UNEMPLOYMENT
. TECHNOLOGICAL UNEMPLOYMENT
. UNDEREMPLOYMENT

UNIVERSITIES AND COLLEGES
. JUNIOR COLLEGES
. MUNICIPAL UNIVERSITIES AND COLLEGES

URBAN ENVIRONMENT
. ARCHITECTURE AND CLIMATE

URBAN PLANNING
. PHYSICAL PLANNING
. URBAN RENEWAL
. . SLUM CLEARANCE
. ZONING

VEHICLES
. AIR CUSHION VEHICLES
. AIRCRAFT
. . BOMBERS
. . HELICOPTERS
. . JET AIRCRAFT
. . SUPERSONIC AIRCRAFT
. ELECTRIC VEHICLES
. . ELECTRIC LOCOMOTIVES
. LOCOMOTIVES
. . DIESEL LOCOMOTIVES
. . ELECTRIC LOCOMOTIVES
. MOTOR VEHICLES
. . AUTOMOBILES
. . . DIPLOMATIC VEHICLES
. . BUSES
. . COMMERCIAL VEHICLES
. . MOTORCYCLES
. . TRACTORS
. . TRUCKS
. PASSENGER VEHICLES
. . PASSENGER SHIPS
. . . SPECIAL TRADE PASSENGER SHIPS
. SHIPS
. . CARGO SHIPS
. . . CONTAINER SHIPS
. . . TANKERS
. . . TRAMP STEAMERS
. . HOSPITAL SHIPS
. . HYDROFOIL CRAFT
. . NUCLEAR SHIPS
. . PASSENGER SHIPS
. . . SPECIAL TRADE PASSENGER SHIPS
. . SUBMARINES
. . . NUCLEAR SUBMARINES
. . WARSHIPS
. . . NUCLEAR SUBMARINES

VETERANS
. DISABLED VETERANS

VIOLENCE
. POLITICAL VIOLENCE
. . RIOTS
. . TERRORISM
. . . HIJACKING OF AIRCRAFT

VOCATIONAL TRAINING
. AGRICULTURAL TRAINING
. . FORESTRY TRAINING
. CO-OPERATIVE EDUCATION
. RETRAINING

WAR AND SOCIETY
. NUCLEAR WARFARE AND SOCIETY

WASTES
. AGRICULTURAL WASTES
. DOMESTIC WASTES
. INDUSTRIAL WASTES
. . SCRAP METALS
. RADIOACTIVE WASTES
. SEWAGE

WATER
. DRINKING WATER
. FRESHWATER
. GROUNDWATER
. . SPRINGS
. HYDROSPHERE
. SALINE WATERS
. . OCEANS
. . . ARCTIC OCEAN
. . . ATLANTIC OCEAN
. . . . NORTH ATLANTIC OCEAN
. . . . . BALTIC SEA
. . . . . BLACK SEA
. . . . . CARIBBEAN SEA
. . . . . MEDITERRANEAN SEA
. . . . . ADRIATIC SEA
. . . . . AEGEAN SEA
. . . . . NORTH SEA
. . . . SOUTH ATLANTIC OCEAN
. . . INDIAN OCEAN
. . . . ARABIAN SEA
. . . . GULF OF ADEN
. . . . PERSIAN GULF
. . . . RED SEA
. . . PACIFIC OCEAN
. . . . NORTH PACIFIC OCEAN
. . . . . EAST CHINA SEA
. . . . SOUTH PACIFIC OCEAN
. . . . . SOUTH CHINA SEA
. SURFACE WATERS
. . LAKES
. . . CASPIAN SEA
. . . GREAT LAKES
. . . LAKE CHAD
. . . LAKE CONSTANCE
. . . LAKE GENEVA
. . . LAKE MALAWI
. . . LAKE TANGANYIKA
. . . LAKE TITICACA
. . . LAKE VICTORIA
. . . MAN-MADE LAKES
. . OCEANS
. . . ARCTIC OCEAN
. . . ATLANTIC OCEAN
. . . . NORTH ATLANTIC OCEAN
. . . . . BALTIC SEA
. . . . . BLACK SEA
. . . . . CARIBBEAN SEA
. . . . . MEDITERRANEAN SEA
. . . . . ADRIATIC SEA
. . . . . AEGEAN SEA
. . . . . NORTH SEA
. . . . SOUTH ATLANTIC OCEAN
. . . INDIAN OCEAN
. . . . ARABIAN SEA
. . . . GULF OF ADEN
. . . . PERSIAN GULF
. . . . RED SEA
. . . PACIFIC OCEAN
. . . . NORTH PACIFIC OCEAN
. . . . . EAST CHINA SEA
. . . . SOUTH PACIFIC OCEAN
. . . . . SOUTH CHINA SEA
. . RESERVOIRS
. . RIVERS
. . . INTERNATIONAL RIVERS
. . . . AMAZON RIVER
. . . . AMUR RIVER
. . . . BRAHMAPUTRA RIVER
. . . . CONGO RIVER
. . . . DANUBE RIVER
. . . . EUPHRATES RIVER
. . . . GANGES RIVER
. . . . INDUS RIVER
. . . . JORDAN RIVER
. . . . MEKONG RIVER
. . . . NIGER RIVER
. . . . NILE RIVER
. . . . ORINOCO RIVER
. . . . PARAGUAY RIVER
. . . . PARANA RIVER
. . . . RHINE RIVER
. . . . RIO GRANDE
. . . . SAINT LAWRENCE SEAWAY
. . . . SALWEEN RIVER
. . . . SHEBELI RIVER
. . . . TIGRIS RIVER
. . . . UBANGI RIVER
. . . . YUKON RIVER
. . . . ZAMBESI RIVER
. . WETLANDS
. . . MANGROVE SWAMPS

WATER MANAGEMENT
. WATER STORAGE
. WATERSHED MANAGEMENT

WEAPON SYSTEMS
. FLEET BALLISTIC MISSILE SYSTEMS
. SMALL CALIBRE WEAPON SYSTEMS
. STRATEGIC NUCLEAR WEAPON SYSTEMS

WESTERN ASIA
. DEMOCRATIC YEMEN
. GULF STATES
. . BAHRAIN
. . KUWAIT
. . QATAR
. . UNITED ARAB EMIRATES
. IRAQ
. JORDAN
. LEBANON
. OMAN
. SAUDI ARABIA
. SYRIAN ARAB REPUBLIC
. YEMEN

WOMEN
. AGING WOMEN
. JEWISH WOMEN
. MARRIED WOMEN
. MUSLIM WOMEN
. PALESTINIAN WOMEN
. WOMEN WORKERS

WORKING TIME ARRANGEMENT
. NIGHT WORK
. OVERTIME
. SHIFT SYSTEMS

WORLD POLITICS
. AFRO-ASIAN POLITICS

YOUTH MOVEMENTS
. STUDENT MOVEMENTS

YOUTH ORGANIZATIONS
. INTERNATIONAL YOUTH ORGANIZATIONS

Exhibit 49

ELECTRICITY
16.09.00 ELECTRICITY

ELECTRIFICATION
05.06.00 ELECTRIFICATION
06.01.01 RURAL ELECTRIFICATION

ELECTROACOUSTICS
05.06.00 ELECTROACOUSTICS

ELECTROCHEMISTRY
03.10.00 ELECTROCHEMISTRY

ELECTRON
16.01.00 ELECTRON MICROSCOPY

ELECTRONIC
14.04.02 ELECTRONIC BEHAVIOUR CONTROL
05.06.00 ELECTRONIC CIRCUITS
05.06.00 ELECTRONIC EQUIPMENT
02.11.00 ELECTRONIC FUNDS TRANSFER
05.06.00 ELECTRONIC INSTRUMENTS
14.04.02 ELECTRONIC SURVEILLANCE

ELECTRONICS
05.06.00 ELECTRONICS
05.06.00 ELECTRONICS INDUSTRY

ELECTRONS
16.09.00 ELECTRONS

ELEMENTARY
16.09.00 ELEMENTARY PARTICLES

ELIGIBLE
18.00.00 ELIGIBLE STATES

ELIMINATION
18.00.00 ELIMINATION

ELITE
14.05.03 ELITE

EMBARGO
01.03.00 ARMS EMBARGO
01.07.04 EMBARGO
01.03.00 GRAIN EMBARGO
01.03.00 OIL EMBARGO

EMBEZZLEMENT
14.04.02 EMBEZZLEMENT

EMBLEMS
14.06.00 EMBLEMS
14.06.00 NATIONAL EMBLEMS

EMBRYOLOGY
10.06.02 EMBRYOLOGY

EMERGENCY
18.00.00 EMERGENCY FUND
01.01.00 EMERGENCY LEGISLATION
10.02.00 EMERGENCY MEDICAL SERVICES
01.01.00 EMERGENCY POWERS
13.02.00 EMERGENCY RELIEF
13.02.00 EMERGENCY SHELTER
18.00.00 EMERGENCY SPECIAL SESSIONS

EMIGRATION
08.01.00 EMIGRATION
01.07.00 EMIGRATION LAW

EMINENT
09.01.00 EMINENT DOMAIN

EMIRATES
08.03.02 OF THE UNITED ARAB EMIRATES
17.05.00 UNITED ARAB EMIRATES

EMISSIONS
06.02.00 AIRCRAFT ENGINE EMISSIONS

EMPLOYEE
12.04.00 EMPLOYEE DISMISSAL
12.03.00 EMPLOYEE MORALE

EMPLOYEES
12.04.00 EMPLOYEES
05.01.02 INVENTIONS BY EMPLOYEES
12.05.00 LOCAL OFFICIALS AND EMPLOYEES
12.03.00 SUPERVISION OF EMPLOYEES

EMPLOYERS
12.04.00 EMPLOYERS

EMPLOYERS'
12.04.00 EMPLOYERS' LIABILITY INSURANCE
12.04.00 EMPLOYERS' ORGANIZATIONS

EMPLOYMENT
12.03.00 AGE AND EMPLOYMENT
12.01.00 EMPLOYMENT
18.00.00 EMPLOYMENT IN ...
12.01.00 EMPLOYMENT POLICY
12.03.00 EMPLOYMENT SECURITY
12.01.00 EMPLOYMENT SERVICES
12.01.00 EMPLOYMENT STATISTICS
12.01.00 FREE CHOICE OF EMPLOYMENT
02.02.00 FULL EMPLOYMENT
12.01.00 NATIONAL EMPLOYMENT SERVICES
12.03.00 PART TIME EMPLOYMENT
12.01.00 RURAL EMPLOYMENT
11.03.00 STUDENT EMPLOYMENT
12.03.00 SUMMER EMPLOYMENT
12.03.00 TEMPORARY EMPLOYMENT

ENCYCLICALS
15.03.00 PAPAL ENCYCLICALS

ENCYCLOPAEDIAS
15.05.00 ENCYCLOPAEDIAS

ENDANGERED
03.07.00 ENDANGERED SPECIES

ENDEMIC
10.03.02 ENDEMIC DISEASES
10.03.02 ENDEMIC GOITRE

ENDERBURY
17.03.02 CANTON AND ENDERBURY ISLANDS

ENDOCRINE
10.03.02 ENDOCRINE DISEASES

ENDOCRINOLOGY
10.06.02 ENDOCRINOLOGY

ENDOWMENTS
14.05.04 ENDOWMENTS

ENERGY
03.10.00 BIOMASS ENERGY
03.10.00 DIRECT ENERGY CONVERSION
03.10.00 ENERGY CONSERVATION
03.10.00 ENERGY CONSUMPTION
03.10.00 ENERGY CONVERSION
03.10.00 ENERGY CRISIS
03.10.00 ENERGY FORECASTS
03.10.00 ENERGY LEGISLATION
03.10.00 ENERGY POLICY
18.00.00 ENERGY REQUIREMENTS
03.10.00 ENERGY RESOURCES
03.10.00 ENERGY STATISTICS
03.10.00 ENERGY TRANSFER
03.10.00 GEOTHERMAL ENERGY
17.07.02 INTERNATIONAL ATOMIC ENERGY AGENCY
03.10.00 MARINE ENERGY RESOURCES
03.10.00 NONCONVENTIONAL ENERGY SOURCES
03.10.00 SOLAR ENERGY
03.10.00 THERMAL ENERGY
03.10.00 TIDAL ENERGY

ENFORCEMENT
14.04.02 LAW ENFORCEMENT
14.04.02 LAW ENFORCEMENT OFFICIALS
14.04.01 NARCOTICS LAW ENFORCEMENT

ENGINE
06.02.00 AIRCRAFT ENGINE EMISSIONS

ENGINEERING
05.05.00 ACOUSTIC ENGINEERING
04.01.02 AGRICULTURAL ENGINEERING
10.06.02 BIOMEDICAL ENGINEERING
05.04.00 CHEMICAL ENGINEERING
05.05.00 CIVIL ENGINEERING
05.06.00 ELECTRICAL ENGINEERING
16.01.00 ENGINEERING
05.05.00 ENGINEERING DESIGN
16.03.00 ENGINEERING GEOLOGY
05.05.00 ENGINEERING INDUSTRIES
15.05.00 ENGINEERING LIBRARIES
16.01.00 ENGINEERING RESEARCH
16.01.00 ENGINEERING SOCIETIES
03.03.00 ENVIRONMENTAL ENGINEERING
04.05.00 FOREST ENGINEERING
10.06.02 GENETIC ENGINEERING
05.06.00 GEOTHERMAL ENGINEERING
06.03.00 HIGHWAY ENGINEERING
16.04.00 HYDRAULIC ENGINEERING
05.03.00 INDUSTRIAL ENGINEERING
06.04.00 MARINE ENGINEERING
05.05.00 MECHANICAL ENGINEERING
12.03.00 METHODS ENGINEERING
05.03.00 MINING ENGINEERING
16.04.00 OCEAN ENGINEERING
05.05.00 OIL RESERVOIR ENGINEERING
05.06.00 PETROLEUM ENGINEERING
05.02.00 PRODUCTION ENGINEERING
06.03.00 RAILWAY ENGINEERING
05.05.00 SYSTEMS ENGINEERING
16.09.00 THERMAL ENGINEERING
06.03.00 TRAFFIC ENGINEERING
06.01.01 TRANSPORT ENGINEERING
05.06.00 WATER SUPPLY ENGINEERING

ENGINEERS
12.05.00 ENGINEERS
12.05.00 WATER SUPPLY ENGINEERS

ENGINES
06.01.02 DIESEL ENGINES
06.04.00 DIESEL MARINE ENGINES
05.03.00 ENGINES
06.04.00 MARINE ENGINES

ENGLISH
15.01.01 BASIC ENGLISH
15.05.00 ENGLISH IMPRINTS
15.01.01 ENGLISH LANGUAGE
15.01.01 ENGLISH LANGUAGE IN FOREIGN COUNTRIES
15.01.02 ENGLISH LITERATURE
15.04.00 ENGLISH NEWSPAPERS
15.04.00 ENGLISH PERIODICALS
15.01.02 ENGLISH POETRY
17.01.00 ENGLISH SPEAKING AFRICA

ENHANCEMENT
16.07.00 PRECIPITATION ENHANCEMENT

ENTENTE
17.07.01 ENTENTE COUNCIL

ENTERIC
10.03.02 ENTERIC DISEASES

ENTERPRISE
05.02.00 BUSINESS ENTERPRISE BRANCHES
02.02.00 ENTERPRISE STATISTICS

ENTERPRISES
05.00.00 BUSINESS ENTERPRISES
05.00.00 INDUSTRIAL ENTERPRISES
05.03.00 MANUFACTURING ENTERPRISES
05.00.00 MINORITY BUSINESS ENTERPRISES
07.04.00 MULTINATIONAL MARKETING ENTERPRISES
05.00.00 PRIVATE ENTERPRISES
02.05.00 PUBLIC ENTERPRISES
05.00.00 SMALL ENTERPRISES
07.01.00 STATE TRADING ENTERPRISES

ENTERTAINMENT
15.06.00 ENTERTAINMENT

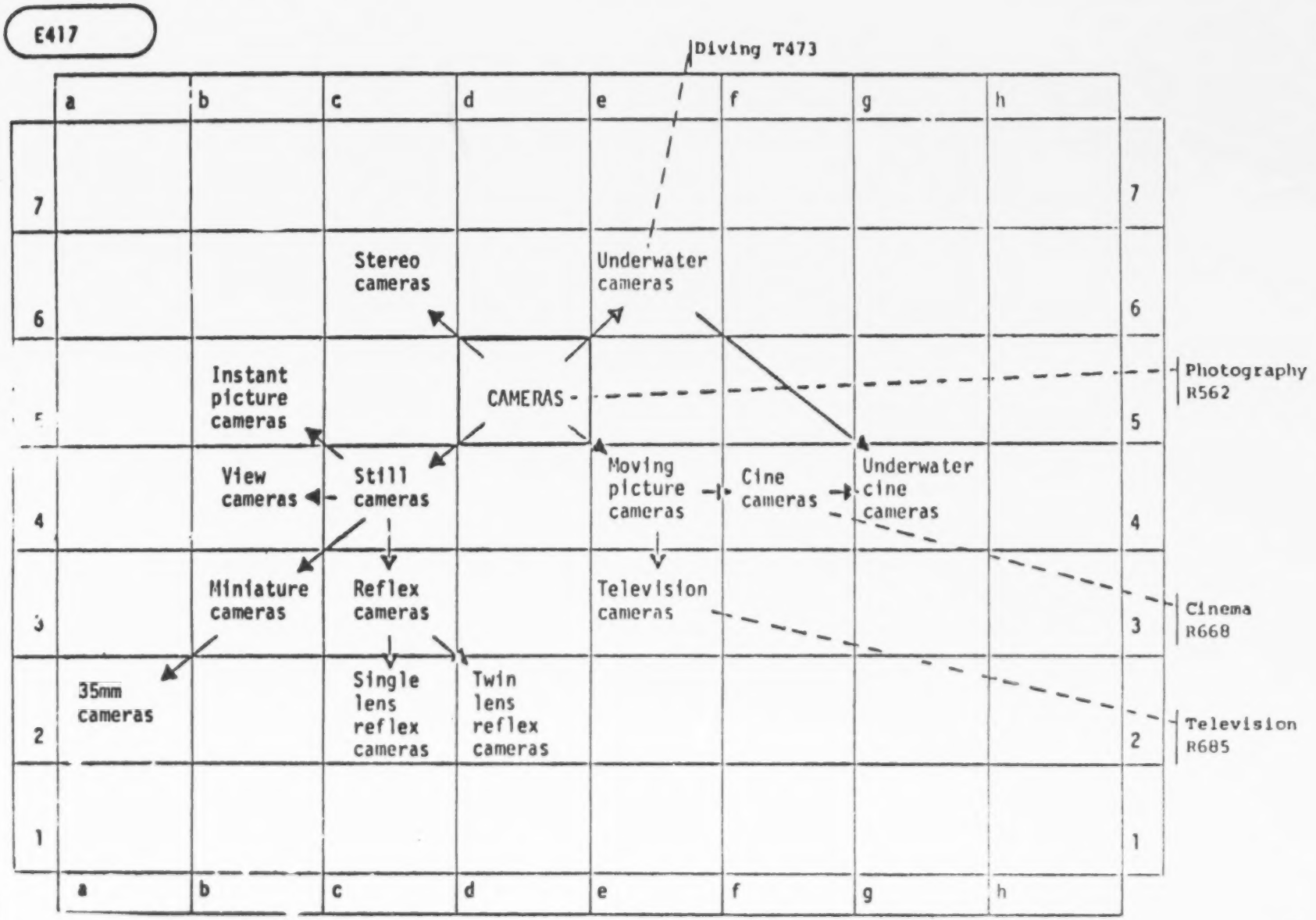
ENTOMOLOGY
16.05.00 ENTOMOLOGY

ENTREPRENEURSHIP
05.02.00 ENTREPRENEURSHIP

ENTRY
15.05.06 CORPORATE ENTRY

ENUMERATION
08.01.06 ENUMERATION

Exhibit 30



74

75

Exhibit 51

35mm CAMERAS E417.a2
BT: Miniature cameras

CAMERAS E417.d5
RT: Photography R562

CINE CAMERAS E417.f4
BT: Moving picture cameras
NT: Underwater cine cameras
RT: Cinema R668

CINEMA R668.d5
RT: Cine cameras E417

DIVING T473.g5
RT: Underwater cameras E417

INSTANT PICTURE CAMERAS E417.b5
SN: Cameras which produce a finished print directly
BT: Still cameras

Land cameras *USE* VIEW CAMERAS

MINIATURE CAMERAS E417.b3
BT: Still cameras
NT: 35mm cameras

MOVING PICTURE CAMERAS E417.e4
BT: Cameras
NT: Cine cameras
Television cameras

PHOTOGRAPHY R562.d5
RT: Cameras E417

REFLEX CAMERAS E417.c3
BT: Still cameras
NT: Single lens reflex cameras
Twin lens reflex cameras

SINGLE LENS REFLEX CAMERAS E417.c2
BT: Reflex cameras

STEREO CAMERAS E417.c6
BT: Cameras

STILL CAMERAS E417.c4
BT: Cameras
NT: Instant picture cameras
Miniature cameras
Reflex cameras
View cameras

TELEVISION R685.d5
RT: Television cameras E417

TELEVISION CAMERAS E417.e3
BT: Moving picture cameras
RT: Television R685

TWIN LENS REFLEX CAMERAS E417.d2
BT: Reflex cameras

UNDERWATER CAMERAS E417.e6
BT: Cameras
NT: Underwater cine cameras
RT: Diving T473

UNDERWATER CINE CAMERAS E417.g4
BT: Cine cameras
Underwater cameras

VIEW CAMERAS E417.b4
SN: Cameras with through-the-lens focusing and a range of movements of the lens plane relative to the film plane
UF: Land cameras
BT: Still cameras

Exhibit 52

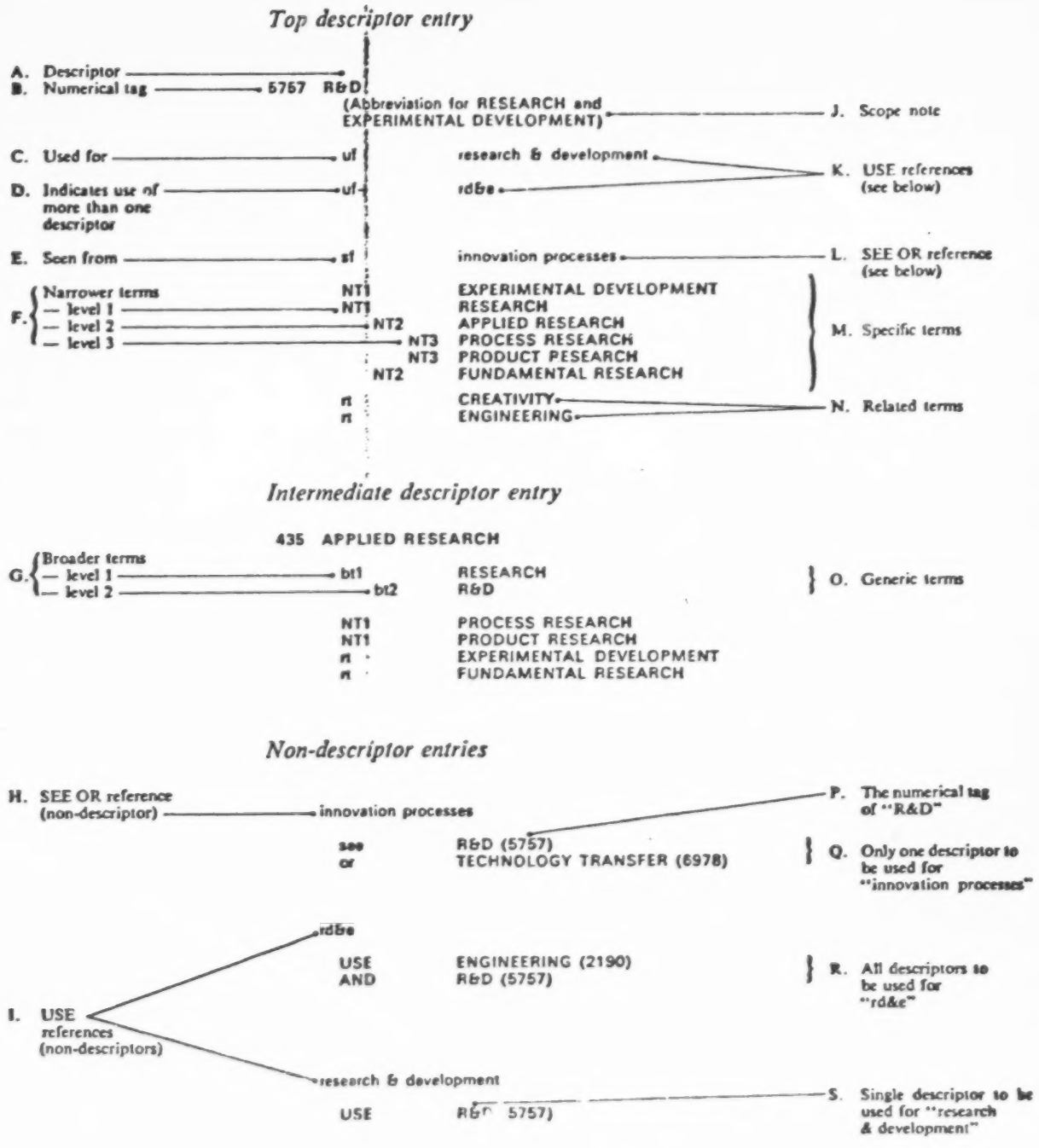

Top descriptor entry
A. Descriptor
B. Numerical tag
5757 R&D
(Abbreviation for RESEARCH and EXPERIMENTAL DEVELOPMENT)
J. Scope note
C. Used for
uf
research & development
D. Indicates use of more than one descriptor
uf
rd&e
K. USE references (see below)
E. Seen from
sf
innovation processes
L. SEE OR reference (see below)
F. Narrower terms
— level 1
— level 2
— level 3
NT1
NT1
NT2
NT3
NT3
NT2
EXPERIMENTAL DEVELOPMENT
RESEARCH
APPLIED RESEARCH
PROCESS RESEARCH
PRODUCT RESEARCH
FUNDAMENTAL RESEARCH
M. Specific terms
rt
rt
CREATIVITY
ENGINEERING
N. Related terms
Intermediate descriptor entry
435 APPLIED RESEARCH
G. Broader terms
— level 1
— level 2
bt1
bt2
RESEARCH
R&D
O. Generic terms
NT1
NT1
rt
rt
PROCESS RESEARCH
PRODUCT RESEARCH
EXPERIMENTAL DEVELOPMENT
FUNDAMENTAL RESEARCH
Non-descriptor entries
H. SEE OR reference (non-descriptor)
innovation processes
see or
R&D (5757)
TECHNOLOGY TRANSFER (6978)
P. The numerical tag of "R&D"
Q. Only one descriptor to be used for "innovation processes"
I. USE references (non-descriptors)
rd&e
USE AND
ENGINEERING (2190)
R&D (5757)
R. All descriptors to be used for "rd&e"
research & development
USE
R&D (5757)
S. Single descriptor to be used for "research & development"

Exhibit 53

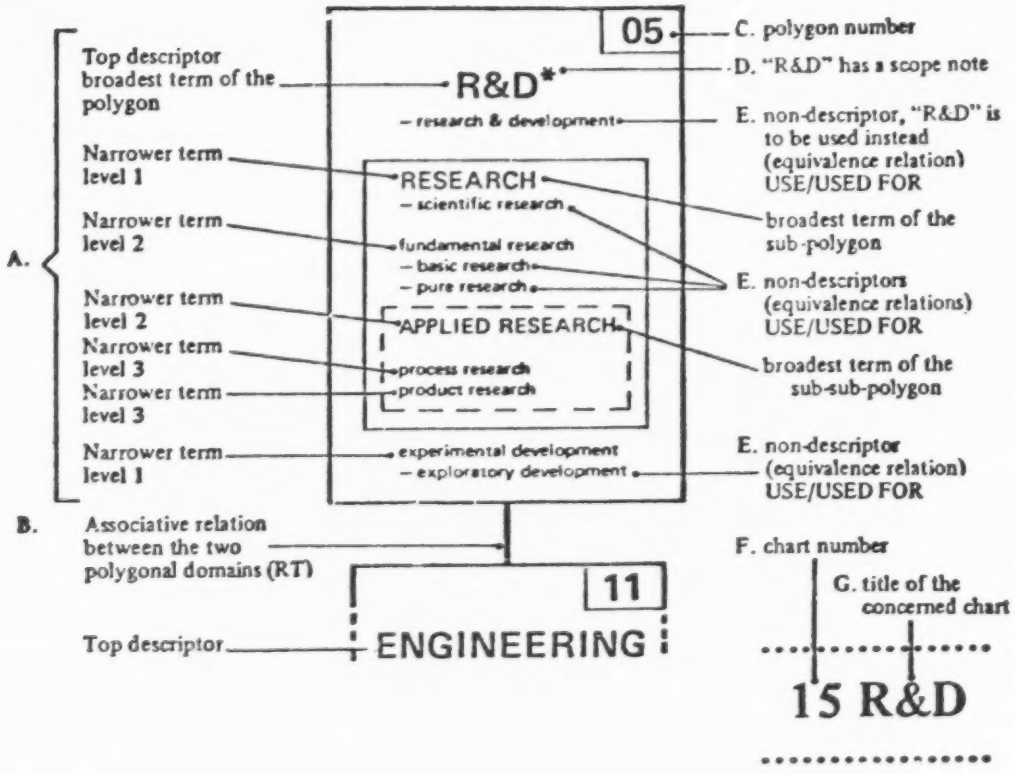


Exhibit 54

Faceted display

La	Acids
Lac	Inorganic acids
Lacd	Hydrochloric acid
Lace	Hydrofluoric acid
Lach	Nitric acid
Laci	Nitrous acid
Lack	Perchloric acid
Lacl	Phosphoric acid
Lacq	Sulfuric acid
Lad	Organic acids

Thesaurus display

Acids	La
RT Corrosive substances	Ps
Sulfuric acid	Lacq
UF Sulphuric acid	
BT(A) Air pollutants	Ppb
RT Acid rain	Pps
Sulphuric acid	
Use Sulfuric acid	Lacq

Exhibit 55

BEST COPY AVAILABLE

R70/84 **Social Problems** (cont.)
R75/78 **Crime**
**Administration of justice M33/39*
**Criminal law M52*
**Law enforcement L56/59*

R75.10 Criminology

R76 Offences
UF Crimes
**Addiction R73.50/90*
**Civil disturbances L73*
**Subversive activities L73.10*
**Terrorism L73.65*
**War crimes M98.50*

(By crimes against property)

R76.10 Theft
UF Burglary
Larceny
Robbery

R76.15 Arson

R76.20 Vandalism
UF Malicious damage

R76.25 Fraud
UF Embezzlement
Forgery

R76.30 Blackmail

R76.35 Hi-jacking

(By crimes against the person)

R76.40 Homicide
UF Murder

R76.40.10 Assassination

R76.45 Violence
UF Assault
Mugging
**Cruelty V15.40.20*

R76.45.10 Domestic violence
UF Violence in the home
Battering
**Ill-treated children R84.60.10*

R76.48 Kidnapping
UF Abduction

R76.50 Hostage-holding

R76.55 Sexual offences
UF Prostitution
White slave trade
Rape

R76.55.40 Pornography

(By crimes against public order)

R76.60 Disorderly conduct
UF Hooliganism

R76.68 Lynching

(By crimes against the state)

R76.70 Political offences
**Subversive activities L73.10*
**Terrorism L73.65*

R76.70.10 Espionage

(By offence against justice)

R76.75 Perjury

R76.78 Corruption
UF Bribery
**Political corruption L76.50*

(By organization)

R76.90 Transnational crime

R76.90.40 Extradition

Craftsmen (cont.)
RT Artists
Handicrafts

Creative role R59.80
BT Role
TT Social structure
RT Creativity

Creative thinking P54.20
BT Thinking
TT Higher mental processes
RT Creativity

Creative writing J46.03 *J60.25 *X60 *Y45.30
SN Students literary composition
UF Student composition
NT Oral composition
BT Literary composition
Study
Writing
TT Communication process
Study
RT Language education
Mother tongue instruction
Study

Creativity P52.10
UF Originality
RT Artistic creation
Creative role
Creative thinking
Cultural creation
Geniuses
Imagination
Visualization

Creches
USE **Day Nurseries**

Credit N88.15
NT Loans
BT Financing
TT Economics
RT Debts
Loans
Mortgages
Subsidies

Credit courses J40.19.13
SN Course work in higher education measured by a credit system whereby credit hours are assigned to a course defined by the number of hours per week in class and the number of weeks in the session.
BT Educational courses
TT Educational courses

Creep C27.30.40C
BT Plastic deformation
TT Damage
Faults
Physics
Properties
RT Fracture
Plasticity

Cremation
USE **Crematoria**

Crematoria E62.30
UF Cremation
RT Cemeteries
Disposal of the dead

Cretaceous period
USE **Mesozoic period**

Crime R75/78 *L57 *M52
NT Offences
BT Social problems
TT Social problems
RT Administration of justice
Crime victims
Criminal law
Criminals
Criminology
Delinquency
Law enforcement
Police
Punishment

Crime victims R84.30
UF Ill-treated persons
Victims of crime
Violence victims
NT Ill-treated children
BT Socially disadvantaged
TT People
Social problems
RT Crime
Cruelty
Violence

Crimes
USE **Offences**

Criminal courts M34.50
BT Courts
TT Courts
RT Criminal law

Criminal investigation L59.50/80
NT Evidence gathering
BT Police activities
TT Police activities
RT Detectives
Forensic medicine

Criminal law M52 *R75
BT Public law
TT Public law
RT Crime
Criminal courts
Criminology
Police

Criminals R77.30
UF Delinquents
Offenders
NT Gangs
War criminals
RT Crime
Criminology
Delinquency
Juvenile delinquency
Maladjusted
Offences
Prisoners

NI/NL **Fluid engineering** (continued)
NJ/NL **Fluid equipment** (continued)
NJQ/NJT **Valves** (continued)

(By component)

NJS.S Valve components
NJS.SB Valve bodies

(By part)

NJS.SBJ Outlet connections
= Gas cylinder connections
* – Gas cylinders NIO.G
[For others synthesize, for example]
NJS.SBS ** Valve bosses
– Valve bodies
+ Bosses

NJS.SJ Valve yokes
NJS.SO Valve plugs
= Plugs (valves)
NJS.SS Tap washers
* – Draw-off taps NJR.D

NJV/NJX **Pipes**
= Tubes (pipes)
*> Delivery pipes (firefighting) GON.D
*> Drainpipes RMF.N
*> Rainwater pipes RJJ.J
* – Bending springs PSS
* – Boiler tubes NEN
* – Bursting discs GVQ
* – Conduits (hydraulic) RCM.C
* – Flues ROC.E
* – Heat engineering NB/NE
* – Hot-water supply systems RMB.M
* – Pressure vessels NIN/NIP
* – Rising mains RMB.FF
* – Small-bore central-heating systems RKK.KR
* – Sewerage GKG
* – Space-heating systems RKJ/RKO
* – Threads NWC/NWD
* – Traps (drainage) RMF.P
* – Ventilation equipment RKR.T

(By property)

NJV.B Rigid pipes
NJV.C Flexible pipes
= Hosepipes
= Hoses
*> Fire hoses GON.B
* – Hose connectors NKB.G
NJV.CC Flexible tubing
= Tubing (flexible)

(By application)

NJV.CP Suction hoses

(By storage device)

NJV.CV Hose reels
*> Fire hose reels GON.BN

> Pipe tees
*– Manifolds NLL.F

Pipe plugs NKD.N
– Pipe couplings
*– Stoppers NSX.F

Pipe reducers
→ Reducing couplings NKD.K

Pipe supports NKH
= Pipe anchors
= Pipe brackets
= Pipe hangers
= Supports (pipes)
< Pipe fittings
– Pipe clips

Pipe systems
→ Pipework systems NLF

Pipe tees NKF.F
= Tees (pipes)
< Pipe junctions

Pipe threads
→ Threads NWC/NWD

Pipe tobacco IYB
< Tobacco
*– Pipes (smokers) XUD

Pipe wrenches PTB.P
< Wrenches

Piped services (laboratory)
→ Laboratory piped services BPE

Pipelaying equipment RUJ
< Construction equipment

Pipeline transportation QBC.Q
< Land transport
*– Pipelines NLC

Pipelines NLC
[Systems of pipes]
= Fluid pipelines
< Fluid equipment
> Asbestos cement pipelines
> Cast-iron pipelines
> Gas pipelines
> Oil pipelines
> Plastic pipelines
– Spigot-and-socket joints
*> Water mains GHS.S
*– District heating RKO.K
*– Drilling rigs STG
*– Flame traps GOJ.J
*– Pipeline transportation QBC.Q

Pipelines
+ Water
=** Water pipelines NLC.R

Piperacese EWK.W
< Dicotyledones

Piperazine DRR.R
= Diethylenediamine
< Heterocyclic compounds

Piperidine DRR.N
< Heterocyclic compounds

Pipes NJV/NJX
= Tubes (pipes)
< Fluid equipment
> Casing pipes
> Flexible pipes
> Gas pipes
> Goosenecks
> Insulation pipes
> Overflow pipes
> Pressure pipes
> Rigid pipes
> Seamed pipes
> Seamless pipes
> Taper pipes
> Vent pipes
– Ducting
– Pipe coatings
*> Delivery pipes (firefighting) GON.D
*> Drainpipes RMF.N
*> Rainwater pipes RJJ.J
*– Bending springs PSS
*– Boiler tubes NEN
*– Bursting discs GVQ
*– Conduits (hydraulic) RCM.C
*– Flues ROC.E
*– Heat engineering NB/NE
*– Hot-water supply systems RMB.M
*– Pressure vessels NIN/NIP
*– Rising mains RMB.FF
*– Sewerage GKG
*– Small-bore central-heating systems RKK.KR
*– Space-heating systems RKJ/RKO
*– Threads NWC/NWD
*– Traps (drainage) RMF.P
*– Ventilation equipment RKR.T

Pipes
+ Asbestos cement
=** Asbestos cement pipes NJW.S

Pipes
+ Ceramics
=** Ceramic pipes VVI

Pipes
+ Plastics
=** Plastic pipes NJW.R

Pipes
+ Water
=** Water pipes NJV.L

Exhibit 59

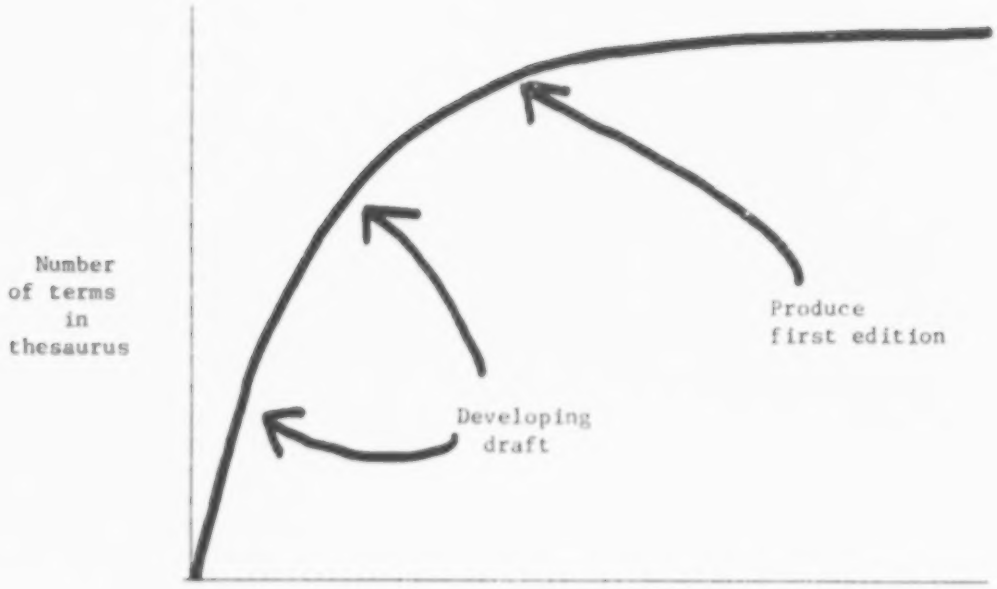

Number of terms in thesaurus
Produce first edition
Developing draft
Number of items indexed (or examined in thesaurus construction)

Exhibit 60

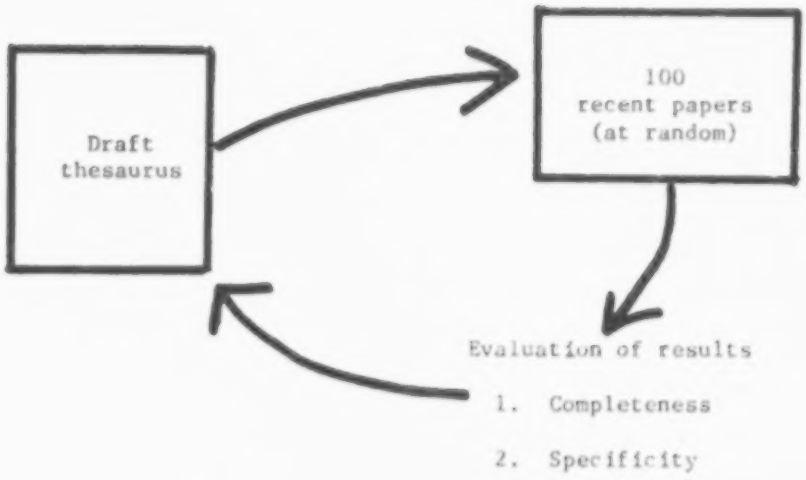


Exhibit 61

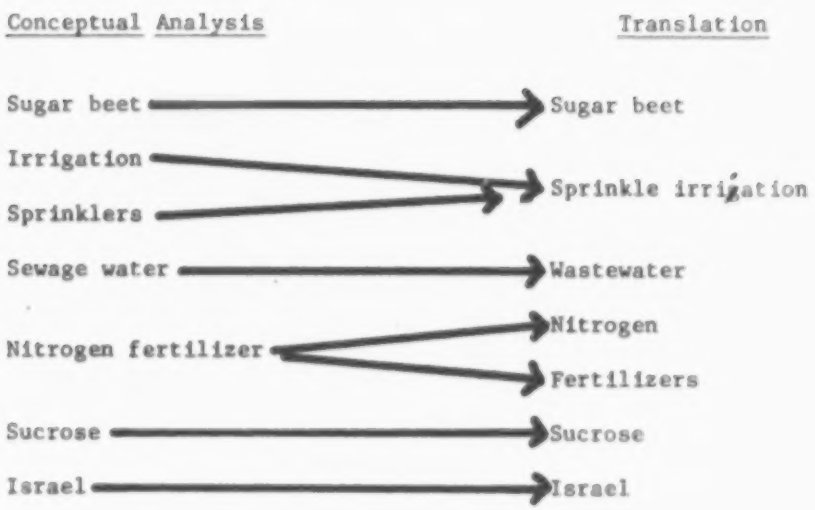


Exhibit 62

Number of concepts	872
Number translatable	860
Percentage translatable	98%
Number translatable at required level of specificity	835
Percentage translatable at required level of specificity	96%

Exhibit 63

Updating

1. Terms from literature
2. Terms from requests

Exhibit 64

Role of the computer

1. Gather terms
2. Expand thesaurus structure
3. Generate printed thesaurus
4. Facilitate updating and maintenance
5. Interface with indexing and searching operations
6. Interface with production of printed index
7. Allow online display

Exhibit 65

Master Vocabulary File

1. Check for consistency and acceptability of terms
2. Allow automatic mapping
3. Maintain statistics on term use
4. Allow searching on complete hierarchies
5. Optimize search strategies
6. Generate cross references for printed indexes
7. Allow online display

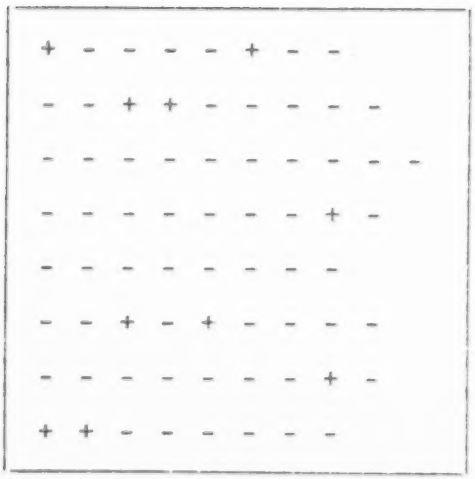
Exhibit 66

1	Flashback	8
2	Flaw detection	207
3	Flaws	82
4	Flexibility	24
5	FLEXIBLE	
6	Flexible filaments	43
7	Flexible walls	58
8	Flexural loading	109
9	Flexural waves	124

Exhibit 67

```
1    CITRUS FRUITS
2         BT   FRUITS
3         NT   GRAPEFRUIT
4              LEMONS
5              LIMES
6              ORANGES
7              TANGERINES
8         RT   CANDIED PEEL
9              FRUIT JUICES
```

Exhibit 68



Search A: +++

Search B: ++++++++
+-------

Exhibit 69

1. Requests that imperfectly represent information needs
2. Indexing factors
3. Search strategy factors
4. Vocabulary factors

Exhibit 70

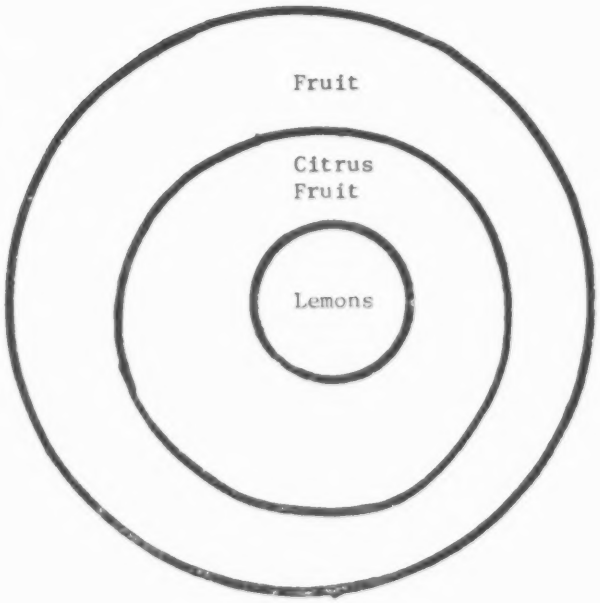


Exhibit 71

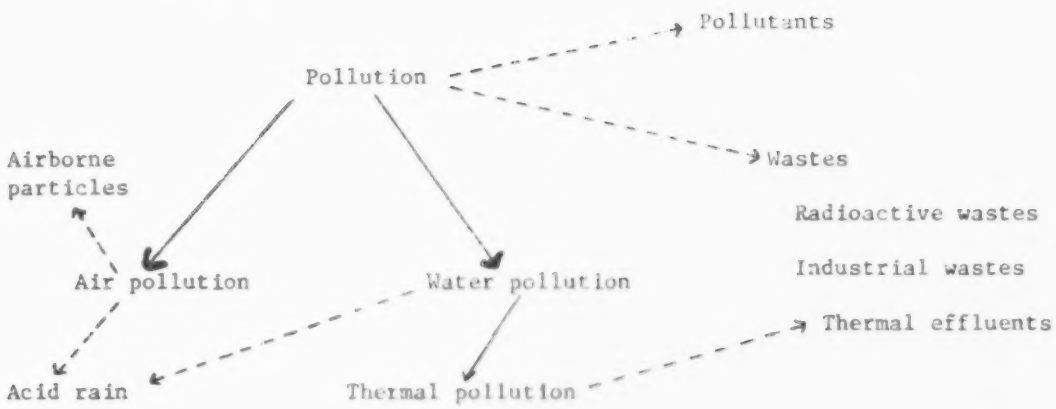


Exhibit 72

BRAZIL

EXPORTS

IMPORTS

BALANCE OF TRADE

ARGENTINA

MEXICO

CURRENCY CONTROLS

EXPORT LICENSES

Exhibit 73

CURRENCY CONTROLS - BRAZIL

BRAZIL - BALANCE OF TRADE

BRAZIL - EXPORTS - ARGENTINA

BRAZIL - EXPORTS - MEXICO

Exhibit 74

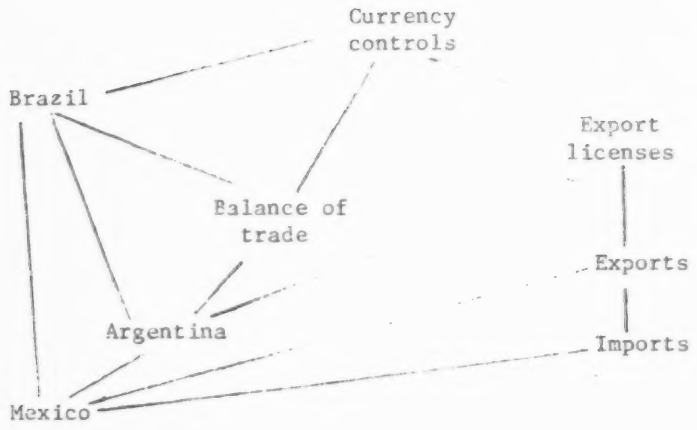


Exhibit 75

ARGENTINA and EXPORTS and MEXICO

BRAZIL) EXPORTS) ARGENTINA) IMPORTS)	1	BRAZIL) BALANCE OF TRADE) CURRENCY CONTROLS) EXPORT LICENSES)	3
BRAZIL) EXPORTS) MEXICO) IMPORTS)	2		

Exhibit 76

A = source, giver

B = target, recipient

Brazil (A)

Export

Argentina (B)

Import

Exhibit 77

Women
Poetry

Women in Poetry

Computers
Design

Computer Design
Airplane Design
Computer-Assisted Design

Computers/Design

Airplanes/Design

Exhibit 78

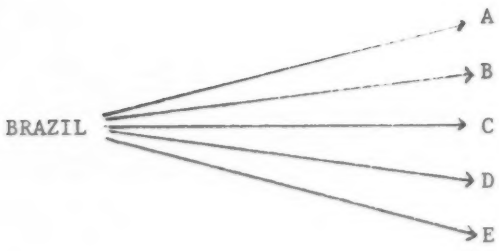


Exhibit 79

WASTES

UF DISCHARGES

EFFLUENTS

SEWAGE

MAGNETS

NT ELECTROMAGNETS

FERROMAGNETICS

RT MAGNETISM

MAGNETOHYDRODYNAMICS

Exhibit 80

Word fragments

Magnet . . .

- Magnet
- Magnets
- Magnetism
- Magnetic
- Magnetically
- Magnetohydrodynamics

. . . magnet

- Electromagnet
- Ferromagnet

. . . magnet . . .

- Electromagnetism
- Ferromagnetism
- Electromagnetically

Exhibit 81

Corrosion . . . . copper

(10 words)

"Copper alloys"

Exhibit 82

LABOR	strikes
WARFARE	strikes

Exhibit 83

WASTE

- WASTE*
- EFFLUENT*
- DISCHARGE*
- SLUDGE*

ANTIBIOTICS

- ANTIBIOTIC*
- *MYCIN
- *ILLIN
- *BIOTIC*
- *MYXIN
- *CYCLINE

Exhibit 84

BIBLIOGRAPHY

Standards

UNISIST guidelines for the establishment and development of monolingual thesauri. Second edition. Paris, Unesco, 1981, 64p. (PGI-81/WS/15)

UNISIST guidelines for the establishment and development of multilingual thesauri. Revised edition. Paris, Unesco, 1980, 85p. (PGI-80/WS/12)

General texts

Aitchison, J.; Gilchrist, A. Thesaurus construction-a practical manual. London, Aslib, 1972, 95p.

Lancaster, F. W. Vocabulary control for information retrieval. Second edition. Arlington, Va., Information Resources Press, 1985 (in press).

Soergel, D. Indexing languages and thesauri: construction and maintenance. Los Angeles, Melville, 1974, 632p. (A useful source of reference rather than a text to be read consecutively.)

Selected journal articles

(These are items dealing with various aspects of the compilation and structure of thesauri that the author has found to be particularly informative or useful.)

Dym, E. D. A new approach to the development of a technical thesaurus. Proceedings of the American Documentation Institute, Vol. 4, 1967, pp. 126-131.

Jones, K. P. Problems associated with the use of compound words in thesauri with special reference to BS 5723: 1979. Journal of documentation, Vol. 37, No. 2, 1981, pp. 53-68.

Mandersloot, W. G. B.; Douglas, E. M. B.; Spicer, N. Thesaurus control--the selection, grouping and cross-referencing of terms for inclusion in a coordinate index word list. Journal of the American Society for Information Science, Vol. 21, No. 1, 1970, pp. 49-57.

Pickford, A. G. A. An objective method for the generation of an information retrieval language. Information scientist, Vol. 2, No. 1, 1968, pp. 17-37.

Willetts, M. An investigation of the nature of the relation between terms in thesauri. Journal of documentation, Vol. 31, No. 3, 1975, pp. 158-184.

"END OF DOCUMENT"